FRR 173

# GIS '92

# WORKING SMARTER

**Vancouver**
**British Columbia**
**Canada**
**February 10 - 13, 1992**

---

**Sixth annual GIS symposium**

---

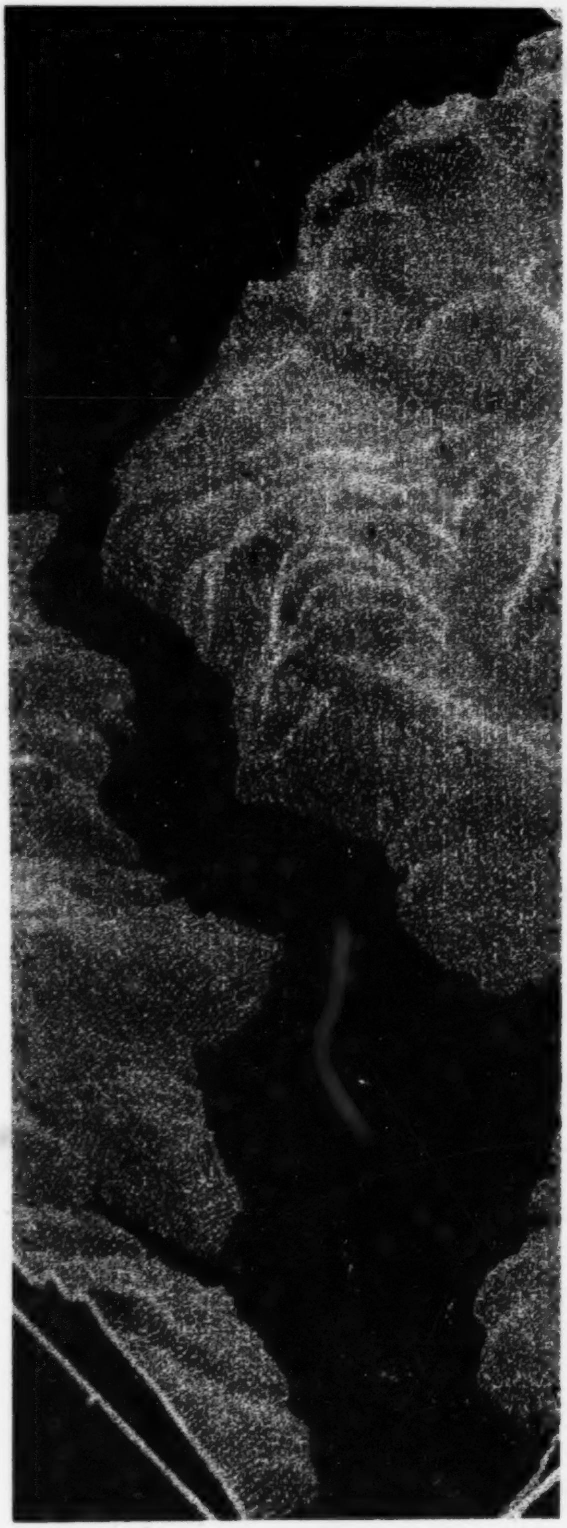
GIS'92 is presented by
Polaris Learning Associates Inc.
and is co-sponsored by
Forestry Canada
and the
Province of British Columbia

---

**polaris***learning*

Forestry Canada Forêt Canada

**Province of**
**British Columbia**
Ministry of Forests
Ministry of Environment,
Lands and Parks



## Symposium Proceedings

G I S

92

Vancouver
British Columbia
Canada
February 10 - 13, 1992

Sixth annual GIS symposium

# WORKING SMARTER

## Symposium Proceedings

*Program Board*

**Paddy O'Reilly**
GIS'92 Symposium Director

**Peter Keller**
Chairman, GIS'92 Technical Program Committee
Department of Geography
University of Victoria

**Frank Hegyi**
Ferihill Technologies Ltd.

**Randy Prokop**
Ministry of Forests

**Don Duffy**
Ministry of Environment, Lands and Parks

*Departmental Representatives*

**Jack de Lestard**
Pacific Forestry Centre
Forestry Canada

**Art Shortreid**
Pacific Forestry Centre
Forestry Canada

FRDA REPORT 173

© Copyright reserved February 1992

polaris*learning*

Forestry Canada
Forêt Canada

Province of British Columbia
Ministry of Forests
Ministry of Environment, Lands and Parks

*This report has been reviewed by Forestry Canada and approved for distribution. Approval does not necessarily signify that the contents reflect the views and policies of Forestry Canada. Mention of trade names or commercial products does not constitute recommendation or endorsement for use.*

103-06947/

# Acknowledgements

**GIS'92,** an *International Symposium on Geographic Information Systems, is sponsored by Forestry Canada, the Province of British Columbia and Polaris Learning Associates, Inc., with funding provided under the Canada-British Columbia Forest Resource Development Agreement.*

**GIS'92 acknowledges additional support from**

Axys Software
Digital Mapping Group Ltd.
Gregory Geoscience Ltd.
IBM Canada
McElhanney Group Limited
PAMAP Technologies Corporation
Sun Microsystems of Canada Inc.
Visitech Graphic Resources Inc.

**GIS'92 is also supported by the following organizations**

Air Canada
Alberta Forestry, Lands and Wildlife
Alberta Research Council
American Society for Photogrammetry
  and Remote Sensing
Applied Science Technologists and Technicians
  of British Columbia
Association of British Columbia Professional Foresters
Association of Ontario Professional Foresters
British Columbia Advanced Systems Institute
British Columbia Forestry Association
British Columbia Institute of Technology
Canadian Cartographic Association
Canadian Institute of Forestry
Canadian International Development Agency
Canadian Pulp and Paper Association
Council of Forest Industries of British Columbia
Forestry Canada
Geomatics Industry Association of Canada
International Union of Forest Research Organizations
Province of British Columbia, Ministry of
  Environment, Lands and Parks
Province of British Columbia, Ministry of Forests
Science Council of British Columbia
Science Council of Canada
Simon Fraser University
University of British Columbia
University of Victoria
Urban and Regional Information Systems Association

*Symposium management and planning by*

**Polaris Learning Associates Inc.**
305 - 1040 Hamilton Street,
Vancouver, BC, V6B 2R9

# Welcome message

International attention has focused on natural resource management. Fiscal expectations have necessitated increased efficiency, and the guiding principle of sustainable development has placed greater pressure on resource management. Geographic Information Systems (GIS) are being used to arrive at solutions to these new demands.

Exploring and promoting the applications of GIS to natural resource management has been the goal of the five major conferences held in Canada, from 1987 to 1992, in which Forestry Canada has participated. This conference is now widely recognized as the world's premier GIS event for forestry and natural resource management applications. I am pleased that Forestry Canada is once again participating, in the 1992 edition of the GIS conference and its presentation of ways to **Work Smarter** with GIS.

GIS applications revolve around people in an organizational context. Its advocates deal with organizational questions such as sponsorship, budget strategies and technical conservatism.

This year's conference continues in the tradition of its predecessors, focusing on GIS applications. The high quality of the GIS symposia have attracted delegates and vendors from increasingly further afield, attesting to the rapidly expanding awareness of the technology, the event, and Canada's preeminence in this area.

I am proud that Forestry Canada continues to champion GIS and encourage greater awareness of the benefits of its technology.

**Honourable Frank Oberle**

*Minister of Forestry*
*Ottawa, Canada*

# GIS'92 Steering and Technical Committee

---

### *Chairmen*

**Norm Smith**
Symposium Co-chair
General Manager
Digital Mapping Group Limited
Burnaby, BC

**Peter Keller**
Chairman
Technical Committee
Department of Geography
University of Victoria
Victoria, BC

### *Members*

**Jack de Lestard**
Forestry Canada
Pacific Forestry Centre
Victoria, BC

**Art Shortreid**
Forestry Canada
Pacific Forestry Centre
Victoria, BC

**Don Duffy**
Surveyor General
Ministry of Environment, Lands and Parks
Victoria, BC

**Frank Edgell**
Ministry of Environment, Lands and Parks
Victoria, BC

**Charles Goodbrand**
GIS Department
Reid, Collins and Associates
Vancouver, BC

**Frank Hegyi**
Ferihill Technologies Ltd.,
Victoria, BC

**Michael Heit**
Ministry of Forests
Victoria, BC

**Munson McKinney**
McElhanney Geosurveys Ltd.,
Vancouver, BC

**Jim McPhalen**
Claymore Consulting Group Ltd.
Vancouver, BC

**Ross Miller**
GIS Program
British Columbia Institute of Technology
Burnaby, BC

**Tom Poiker**
Department of Geography
Simon Fraser University
Burnaby, BC

**Randy Prokop**
Ministry of Forests
Victoria, BC

**Hans Schreier**
Department of Soil Science
University of British Columbia
Vancouver, BC

**Norm Shaw**
Forest Resource Program
British Columbia Institute of Technology
Burnaby, BC

### *GIS'92 Resource Staff*

**Paddy O'Reilly**
Symposium Director

**Ihor Zalubniak**
Symposium Manager

**Dina Grskovich**
Administrative Assistant

**Brenda Corby**
Registration Coordinator

**Gabriole Springford**
Desktop Publishing

**Donal O'Reilly**
Proceedings Coordinator

**Tempest Deptuch**
CIDA Group Coordinator

*This publication contains the text of papers presented at GIS'92 that were received prior to January 26, 1992. It is being made available for the information of participants and to stimulate discussion during the Symposium. Where an author has been unable to meet the deadline for this publication, an abstract has been included.*

*Papers were produced from disk, hard copy and graphics prepared by authors and no responsibility whatsoever is accepted by the organizers of GIS'92 for any errors, omissions or breach of copyright in their content. Papers are grouped according to the sessions in which they were presented.*

## Symposium themes

**A. The corporate database**

Implementation for smarter working
Access to data: grief or relief?
Data integration: pay up or pay-off?

**B. Applications of GIS in forest management**

Inventory and planning
Visual impact modelling
Towards the year 2000 — FORMAN
Fire applications of GIS
Harvesting and operations

**C. Environmental applications of GIS**

Wildlife modelling
Habitat modelling
Environmental information management
Environmental information analysis — case studies

**D. GIS implementation and techniques**

Implementation and impediments
Evolving methodologies in GIS
Data issues
Scanning and conversion
Integrating images with GIS

**Special interest group**

Human Resources Working Group

**Plenary presentations**

A digital elevation model of the GIS navel
Working smarter amidst chaos
Smarter applications — where do we go from here?
New technologies for creating public windows into GISs
Environmentally smart working — fact or fancy?
Are we really working smarter in our forests — or just faster?
Technologies and methodologies — are better solutions emerging?
The corporate data base — making the marriage work

ISSN 0835-0752

ISBN 0-662-19392X

Ministry of Supply and Services Canada, 1992
Cat. No. FO 18-16/1992E

*Editing, copyfitting and desktop publishing by*
*Polaris Learning Associates Inc., Vancouver, B.C.*

*Printing by Gastown Printers, Vancouver, B.C.*

# Contents

## A The corporate data base

### A1 Implementing for smarter working

Designing British Columbia's land information infrastructure: from vision to reality *Evert Kenk*

Efficient land information use within the province of British Columbia: Policy development based on need *Robert Menes*

Design and operation of a meta-GIS *Rick Vrana*

The use of GIS as a dynamic multiple-use management tool by the U. S. Bureau of Land Management on federal public lands in Colorado *Bob Moore*

### A2 Access to data: grief or relief?

The GRASS CD-ROM global data base and public domain GIS software *Scott Madry*

Access to data in support of local studies and global investigations *John E. Estes*

Geographic information system for multidata database management *Frank Indome*

Spatial information sharing spells grief or relief ? *Diane Allen*

### A3 Corporate data integration

Integrating ecological and forestry databases using GIS *Mike Comeau*

Developing integrated/interfaced geographic information systems *Yanni Xaio*

Linking a spatially-oriented Northern Spotted Owl population dynamics model with the Western Oregon Digital Database *Duane Dippon*

Multi-ownership integration of forest stand data: A case study in Itasca County, Minnesota *Greg Arthaud*

Alberta's vegetation information system *Richard Haag*

---

*This list of contents shows paper title and name of presenting author only. Full credits appear on the papers. To allow for last-minute production, we have not sequentially numbered pages in these Proceedings. To locate a paper, first find the appropriate session, then browse until you find the paper you are looking for.*

# B Applications of GIS in forest management

## B1 Inventory and planning

Approaches to forest survey, inventory, evaluation and planning using biophysical models of landscape processes *T.W. Norton*

Forest monitoring using temporal GIS techniques *Gail Langran*

Analysis of the spatial distribution of a forest population for the planning of a forest inventory *Jean-Marie Bilodeau*

Quality control in data base development and spatial analysis *Brendan Mackay*

## B2 Visual impact modelling

Viewscape analysis: a comprehensive methodology *Charles Goodbrand*

Design and simulations of harvesting by visual quality objectives class *Catherine Berris*

Decision support tool for wood harvesting mangement in regard to landscape quality integrity *Olivier Richer*

GIS for visual impact assessment – a case study from the Three Sisters Resort development, Alberta *David Buckley*

## B3 Towards the year 2000: FORMAN

GIS and forest management — towards the next century *Glen Jordan*

GIS and forest management planning — improving operational reality in spatial modelling *Boyd Pittman*

GIS and forest management planning — integrating strategic and operations considerations *David Carleton*

Up the GIS learning curve — the New Brunswick forest management experience *Bob Dick*

## B4 Fire applications

Tracking suppression resources via an Australian fire management system: in principle and in practice *Stephen Kessell*

Evaluating spatial strategies of wildfire prevention with a GIS *Yue Hong Chou*

The application of GIS technology for forest fire management planning *Bryan Lee*

## B5 Harvesting operations

Integration of silvicultural record keeping and geographic information systems *Robert Kennett*

GIS-based strategic planning at MacMillan Bloedel Limited *Brad Whitehead*

Elements for tracking forest planning in Quebec with the aid of GIS *Gilles Arseneau*

Probability based spatial-temporal modelling for natural forests: sampling techniques, model development, accuracy assessment, and residual evaluation *Kim Lowell*

# C Environmental applications of GIS

## C1 GIS for wildlife modelling

The Map Machine: using the structured query language to produce precisely tailored GIS maps of telemetry observations *Norm Cimon*

Integrating spatial models of recreation, wildlife, and timber resources *Lloyd Queen*

A biophysical process approach to wildlife modelling in forestry *T. W. Norton*

BC Indian fisheries information system communication techniques for consultation and negotiation *Joe Kuhn*

## C2 GIS for habitat modelling

Continuous habitat suitability modelling using convolution filters *John Rickers*

The Ecosystem Management Model project: integrating simulation modelling and GIS in the Canadian Parks Service *David Buckley*

Using GIS to analyze Spotted Owl habitat within Oregon *Pat Green*

Managing and monitoring wildlife habitat in the North Yukon *poster* *Manon Desforges*

## C3 GIS in information management

Mapping and establishing GIS for tropical rain forest in Indonesia *Herman Hidyat*

Reliability of GIS for environmental assessement in data poor regions *Kevin Barthel*

Using GIS and image processing to prioritize cumulative effects analysis *Steven Bernath*

An examination of the role of GIS in the regulatory process: the Oslo project as an example *James Case*

Monitoring and modelling transboundary air quality using a GIS *Anne Lucas*

## C4 GIS in information analysis

Fast access to fisheries information with low-cost desktop mapping software *Peter Lewis*

The conceptual determination of carrying capacity analysis using GIS technology. Case study: the upper Komering sub-watershed, South Sumatera, Indonesia *Nanna Suryana*

Setting global and regional priorities for coastal temperate rain forest conservation *Andrew Mitchell*

Evaluating alternative land management policies from a water quality perspective: an automated land information systems approach *Lucy Buchan*

# D GIS Implementation and techniques

---

## D1 Implementation and Impediments

| | |
|---|---|
| GIS two years after purchase (Where are we?) | *Craig Sharun* |
| Distributing GIS data in a computer network | *Bert Veenendaal* |
| MIS and GIS - can they work together? | *Bruce Abbott* |
| Trends in GIS technology. What are the impediments? | *Carl Reed* |

## D2 Evolving methodologies

| | |
|---|---|
| The use of stream drainage patterns in a GIS program package for the delineation of geologic faults and fractures | *Richard Thiessen* |
| A decision support system using neural nets and genetic algorithms for multi-objective State Park management | *Randy Gimblett* |
| Parser design strategies for GIS natural language interfaces | *William Lydiatt* |
| Phase II EPPL shell: A GIS application for the field resource manager | *Dave Martodam* |

## D3 Evolving methodologies

| | |
|---|---|
| A corporate database approach to GIS: techniques for integrating multiple GIS applications into the corporate data environment | *John Herring* |
| Resource management in the 1990s using statistical and spatial modelling | *Stephen Sperry* |
| Integrating spatial analysis and GIS through application program interfaces (AIP) | *James Strapp* |
| Inexact spatial reasoning | *David Altman* |

## D4 Evolving methodologies

| | |
|---|---|
| Resource mapping and traversing with the Global Positioning System in northern and central British Columbia | *Jamie McLennan* |
| Creating accurate GIS databases with digital photogrammetry | *Stephen Sperry* |
| An advanced software engineering environment for prototyping and development of GIS applications | *John Smyrnew* |
| Analysis of GIS user requirements within the Yukon Department of Renewable Resources using CASE tools | *William Johnston* |

## D5 Data issues

| | |
|---|---|
| SAIF - conquering space and time | *H.A. Kucera* |
| NAD27 to NAD83 conversion in British Columbia: National Transformation, Version 1.1 | *Amin Kassam* |
| Putting the records in order: GIS database construction | *C. Warren Ferguson* |
| Technology trends for GIS applications | *Michael J. Fisher* |

## D6 Scanning and conversion

Automated map data conversion *Wayne Coleman*
Impacts of scanning resolution on GIS data conversion *Lisa Klapwyk Crouser*
Scanning technology: considerations in the preparation of scanner-ready manuscripts *Daniel Haskell*
Data collection for GIS applications using scanning in conjunction with automated raster to vector conversion technology *Jeff Marceil*

## D7 Integrating images with GIS

Utilizing satellite data and GIS to map land cover change *Paul Maus*
Replacing aerial photo interpretation in resource inventories with integrated digital images and GIS *Frank Hegyi*
Managing forest resources with raster tools *Andrew Bury*
Using ancillary data in post-classification modelling to increase the accuracy of conifer species classifications from LANDSATÔ data *Mike Golden*

# Human Resources Working Group

Hardware, software and flesh: a human resources architecture for geographic information systems *Robert Menes*
New GIS/LIS education and training initiatives in Australia and southeastern Asia *Stephen R. Kessell*

# Reserve and poster papers

GIS management of felled and bucked log inventories susceptible to Ambrosia beetle attack *Rick Howe*
Recent developments in spatial analysis of fisheries acoustic data *R. Kieser*
The next generation of digital spatial technologies: error utilization, thematic map and remote sensing integration, spatial operators and data structures *Kim Lowell*
Spatial considerations for forest canopy modelling and biomass estimation based on remotely sensed data *Geoffrey Edwards*
Regeneration stocking derived from high resolution airborne imagery via texture measurements *Geoffrey Edwards*
An expert system for helping forest managers choose among data sources of remotely sensed imagery *Geoffrey Edwards*
A geographic information system for strategic and operational monitoring of forest activity: a conceptual approach *Riadika Mastra*
Eco-camp: and environmental education experience *Timothy Day*

# Plenary presentations

| | |
|---|---|
| A digital elevation model of the GIS navel | *John Keene* |
| Working smarter amids chaos | *Hans Schreier* |
| Smarter applications — where do we go from here? | *Joseph K. Berry* |
| New technologies for creating public windows into GISs | *Jack Dangermond* |
| * Environmentally smarter working — fact or fancy? | *Ellsworth LeDrew* |
| * Are we really working smarter in our forests — or just faster? | *David Goodenough* |
| * Technologies and methodologies — are better solutions emerging? | *Michael F. Foodchild* |
| * The corporate database — making the marriage work | *Stephen Guptill* |

** Abstracts not available at time of printing*

# A1 Implementing for smarter working

---

**Designing British Columbia's land information infrastructure: from vision to reality**

Evert Kenk et al., Ministry of Lands and Parks, Victoria, BC, Canada

**Efficient land information use within the province of British Columbia: Policy development based on need**

Robert Menes et al., Ministry of Environment Lands and Parks, Victoria, BC, Canada

**Design and operation of a meta-GIS**

Rick Vrana, Portland State University, Portland, OR, USA

**The use of GIS as a dynamic multiple-use management tool by the U.S. Bureau of Land Management on federal public lands in Colorado**

Bob Moore, Bureau of Land Management, Lakewood, CO, USA

---

A1 Implementing for smarter working

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Designing British Columbia's land information infrastructure: from vision to reality

**Evert Kenk**

Head, Government Land Information Data Exchange
Surveys and Resource Mapping Branch
BC Environment, Lands and Parks
4th Floor, 1802 Douglas St.
Victoria, B.C. V8V 1X4, Canada
Tel: 604-387-9309
Fax: 604-356-7831

**Tony Baker**

Project Engineer
MacDonald Dettwiler
13800 Commerce Parkway
Richmond, B.C. V6V 2J3, Canada
Tel: 604-278-3411
Fax: 604-278-2117

## Abstract

*The Government of British Columbia has recognized the importance of exchanging and sharing its land-related data both internally and with outside client groups. To this end, it has created the Corporate Land Information Strategic Plan (CLISP) to enable the integration of diverse government land information systems while maintaining the autonomy of the individual ministries in managing those information systems.*

*As a key part of the CLISP, the Ministry of Environment Lands and Parks has been given the task of establishing a land-related data access and management system called the Land Information Infrastructure (LII). The LII will be developed in stages over the next seven years and will involve the active participation of several key ministries and other government agencies that are custodians, providers and users of land-related data.*

*A vision of the function and architecture of the target LII system in 1998 has been defined. The target system will provide prompt, consistent and flexible access to land-related data maintained throughout the B.C Government as well as a means for exporting the data to user application systems. The system will use a versatile heterogenous multi-database architecture.*

*A migration strategy has been prepared that charts a development path from the current situation to the target system. It takes a phased approach that will provide early benefits to users while minimizing development risk. The strategy includes a plan for the involvement of the participating groups and a description of intermediate implementations of the LII system.*

## Introduction

The Government of British Columbia is a large, diverse organization. Like many other government jurisdictions it manages large amounts of dispersed, land-related spatial data. Through the 1980's, with the rapid development of land information systems, various government agencies automated their existing major data collection programs to more effectively deliver their mandates. These developments were undertaken as independent, single business purpose initiatives resulting in a heterogenous, multi-database environment which includes differences in hardware, operating systems, database architectures and database management systems. Although many of these systems meet the business needs of the custodian agencies for which they were first built, they all contribute to a common problem that must be solved: how to provide efficient access, and effectively exchange and share land-related, spatial data from these many disparate sources.

In 1989, the Government of British Columbia adopted the Corporate Land Information Strategic Plan (CLISP) to enable access to, and facilitate the sharing and exchange of the diverse land-related data holdings in Government. CLISP is a strategic plan for managing land-related data and systems in the B.C Government. It is corporate, extending across ministerial boundaries. It is also strategic providing both a target vision for land-information and supporting systems and an overall strategy for achieving that vision.

The implementation of CLISP is being overseen by an inter-ministerial Land Information Strategic Committee (LISC) chaired by the Ministry of Environment Lands and Parks. The Ministry of Environment Lands and Parks was mandated as the central agency for land information systems and has the responsibility to create and coordinate standards for the management of land-related information and systems with a focus on enabling access to and facilitating the exchange and sharing of data between government agencies. The Government Land Information Data Exchange (GLIDE) Unit was created to manage CLISP and to provide information on its progress and operation.

CLISP has three main aspects: organization, policy and system. Each aspect involves the development of plans and specific tasks to meet the B.C. objective of servicing the land information needs of government as a whole in a standard, transparent and convenient manner. This paper focuses on one aspect; system. For more information on organization and policy refer to (Hofmeyr et. al., 1991).

The central deliverable of CLISP is a physical infrastructure system including computer hardware, software, communications and standards, referred to as the Land Information Infrastructure (LII). The LII is being developed within a corresponding policy framework referred to as the Land Information Management Framework (LIMF). The LII will provide the technology and information required to enable the sharing of land-related information among the various agencies of the Provincial Government.

## The vision

The past decade has made us more aware of our finite resources, the costs associated with managing them effectively and we have come to realize the extreme interdependency of our actions and the numerous physical and biological processes which govern the world around us. Those responsible for managing and developing our resources are being forced to manage more intensively and develop new skills requiring better use of available information. Planning and management will become more geographically specific. Each and every land or resource management issue will be resolved with a wide variety of information. Applications for supporting the business activities of government such as the management of timber supply, maintenance of land status, oil-spill response and other complex decision support will be developed. To be successful all such applications will require access to a wide variety of land-related information. Currently this is not possible, however because it is beyond the scope of any individual system or individual to deal with the problems of "global" data access and data translation.

To this end, the vision of the LII of the future, 1998, is a system that efficiently and effectively meets the needs of its users by providing easy, prompt, consistent and flexible access to land-related data, irrespective of where the user is located or where and how the data are stored and to have that data presented in a form most beneficial to the user and/or application to which it will be applied, Figure 1. The target system will serve both groups within and outside the B.C. Government, including private corporations, other levels of government and the public.

The LII will provide its users with a single window into the land-related information sources in Government and not compromise the security and integrity of the data it accesses. Ministry custodians of data sources will retain autonomy in managing their databases. Users or user applications will gain access to the LII via User Access Terminals (UAT's), which are software components running on their local hardware platform. The UAT provides the means, via access to LII services and the required information about data sources (contained in the LII Repository), by which the user can find out what land-related data is accessible using the LII, to order data, and to receive the retrieved data. The UAT will enable users to join, superimpose and manipulate data from directly linked (on-line) disparate digital data sources and deliver it for use on their local system or application program with minimum effort. The user will also be able to formulate requests for data from off-line data sources not directly linked to the LII. In this latter

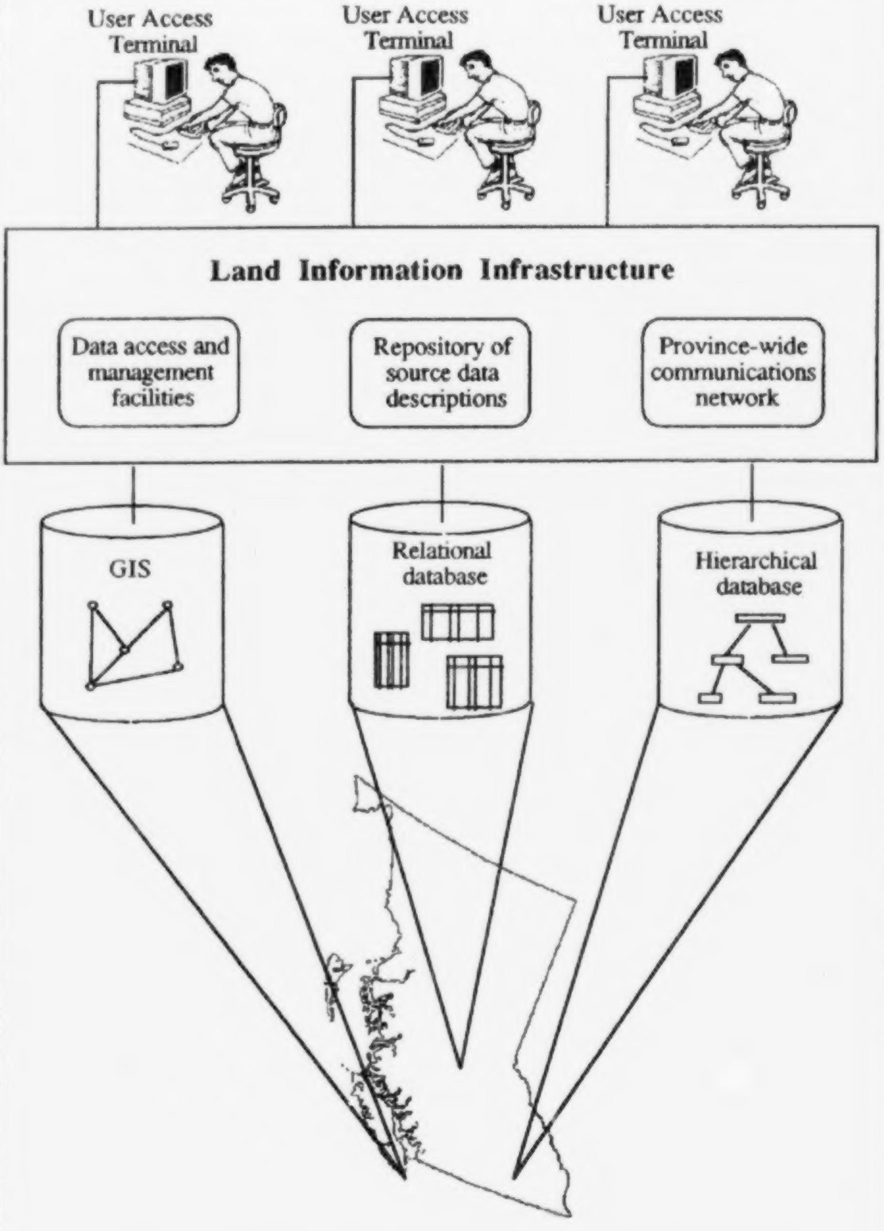

User Access Terminal
User Access Terminal
User Access Terminal
Land Information Infrastructure
Data access and management facilities
Repository of source data descriptions
Province-wide communications network
GIS
Relational database
Hierarchical database

Figure 1. The Land Information Infrastructure Concept

case, the LII dispatches the request to the custodian of the off-line source who then physically fills the order. This may be for hardcopy products not available in digital form. The user will formulate his requests using a rich text and graphics interface provided by the UAT that will support queries by forms, graphics or example.

Because the LII mediates between data sources and users, and undertakes the management of data access and delivery, in addition to data linking from disparate data sources, it will appear as if the user is ordering data from one, or a few, databases, each of which in reality is an integrated view of more than one database. For example, there are at least five databases in Government which show legal encumbrances on the land. In the vision, the LII would provide a single integrated view to these databases. If the user, via his/her UAT were to define a geographic area and queried the LII for all legal encumbrances within that area, the user, to this single query, would receive a composite of consistent data, derived from all the databases, in a form suited for use on his/her application. To the user, the data would have appeared to have come from one database.

The LII will enable the valuing (and pricing, if desired) of land-related data by recording transactions and so allowing those who capture source data to understand by who, how, and to what extent, the data is being used. The target system will provide distributed, automated support to allow data and database mangers and administrators to maintain the information the LII requires about their data and systems, check data integrity and to report problems. The system will also report on its performance and health to operations personnel.

Systems support to the business environment in which the LII will operate in 1998, will be distinctly different than it is now. It is likely that the trend towards vertical integration of business systems will continue with the emergence and maturity of "enterprise or corporate" systems. These are systems in which all applications, at all levels within an organization, supporting a particular business area are managed and integrated in one system. Such business systems will evolve in the British Columbia Government spanning entire ministries and even cross ministerial boundaries. These systems will support day-to-day transaction activities, provide statistics and reports to regional and central managers and put key geographic information derived from land-related databases at the fingertips of decision makers. Examples to illustrate include an Environmental Management System supporting the vetting, issue and management of all types of Environment tenures; management of the processing of referrals; wildlife inventory, habitat and suitability monitoring; and reporting at all levels. A similar Integrated Forest Management system could evolve supporting the management of all types of forestry tenures; forest inventory management; forest growth, yield and harvest modelling; and reporting at all levels of management from operational to consolidated reports to the Chief Forester. The LII although not an enterprise system, as defined here, will play a major role by providing the means of exchanging timely, specific and consistent land-related information within and between such enterprise systems.

Although the LII will provide GIS-like functionality to facilitate the ordering of data in the context of a spatial (e.g. graphic, map, or geographic) view, it should not be confused with providing a corporate-GIS meeting the many and varied ministry specific business requirements. The LII is not a forest management system, nor is it an environmental monitoring system, nor is it a system for analyzing competing land use policies. The LII is the system whereby the land-related data resources of the Province can be made available to and shared among such systems.

## Charting a course

The previous section described the vision from a users point of view of the target system. User expectation is difficult to manage and it must be realized that the target system is not intended to be fully operational until 1998. Figure 2 shows the timeline for the phased development of the LII. The phased migration strategy must balance the needs of users against the reality of limited time, resources, technical limitations of existing data sources and government communication networks, and the readiness of organizations to make effective use of data provided by the LII. With that in mind, a phased migration strategy has been charted which provides early benefits to users, provides increased functionality with each implementation to match expanding requirements and to minimize development risk.

Phase I is complete. Its output was a conceptual design for the Phase III system and the high-level design for the Phase II operational prototype system.

At this time three major versions of the LII are envisaged before emergence of the target system in 1998. The Phase II prototype system will be operational in early 1993. It will be a limited implementation intended to be both useful, but also a platform from which the first full system will be developed. It will enable access to a small number of on-line databases and a wider set of off-line databases. The distribution of UATs will be limited to approximately a dozen key locations in participating ministries. The intent is develop a system which provides useful service to a sizeable number of users to demonstrate the capability of the system and to demonstrate to its developers and custodians that the design and architecture are capable of being extended into a Phase III production system.

Requests will be formulated using a Common Data Manipulation Language as graphical formulation of requests will be rudimentary. Most on-line data will be

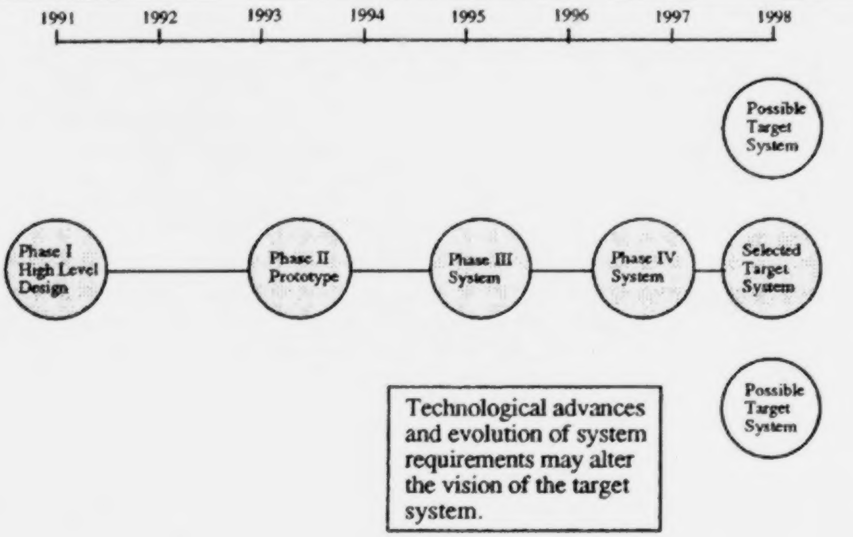


**Figure 2. The Migration Strategy Roadmap**

delivered electronically and it will be possible to display graphically, spatial data delivered in the LII's common data interchange format, but a custom translator may have to be supplied by the user to put the data in a form useable by his/her application. The system will have full transactions and resource accounting, standard security and access control, on-line help and support for the LII Data Administrator to update the LII Repository.

The Phase III system will be operational in early 1995. It will allow access to more on-line databases and will permit coherent retrieval and joining of data from multiple on-line databases. The Phase III system will be a true transaction processing integrated multi-database system. It will be able to support in the order of 100 concurrent users. Users will be able to use a comprehensive interactive graphics-based method for defining the data they require. For example a user could start with a map of B.C., draw an arbitrary closed shape and then request data from a number of sources falling within that boundary. Spatially-based queries will be supported. For example a user could identify an area displayed on a topographic base map and ask which Crown tenures apply to it. Users will have personal storage space within the LII allowing them to build their own body of information over time to facilitate their use of the LII.

The Phase IV system, which will be operational early in 1997 will be an extension to the Phase III system and will be more easily extensible and database update may be possible.

The ultimate success of this migration strategy depends on the close cooperation with all participating groups. Participating ministries must be willing and able to commit the resources to make their key land-related data available for access via the LII. Ministries must be willing to make changes over time to effectively use data provided by the LII. The hardware, communications and software technology, and its management, must be in place to support use of the LII. The custodians of the LII must provide the management, liaison, consulting, training and other services necessary for the day-to day use of the LII.

## Designing for success

### System requirements

The LII is a system that provides transparent integrated access to sources of land-related data. For the most part, these sources are databases on digital computers connected on-line to the LII via communications networks. Users of the LII are able to find out what land-related data are available for access, and then request selected spatial and attribute data without having to know where or how the data are stored (see Figure 1). For example, someone might require the names of lessees of Crown tenures that border a given road. A query could be issued to the LII to obtain this information assuming it is accessible via the LII. To fulfil this query, the LII might need to retrieve and process Crown tenure polygon data from a GIS, road segment data from a digital map file system and attribute data from a relational DBMS, each of these being separate systems in different locations. Spatial data are transferred and manipulated within the LII in the Spatial Archive and

Interchange Format (SAIF), the draft national standard for interchange of geomatics data.

These are the key requirements of the LII:

*Extensibility:* The LII will evolve in stages. It is important that connecting new source databases does not require redevelopment of the existing system.

*Flexibility:* The LII will have an operational life of many years and will serve the needs of a large and diverse organization, namely the B.C. Government. The system must be flexible so that changes in its operating environment (e.g. changes in organizational structure, physical locations, policy, issues of concern) can easily be accommodated.

*Expandability:* When the LII becomes operational, the demand for its services is likely to be modest initially, but will increase substantially as its benefits are realised and as new databases are connected. Also, there are likely to be requests for additional services. The LII design must be such that system performance can be increased and the services it provides can be expanded to meet rising demand.

The design for the LII addresses these and other requirements and is based on a federated heterogeneous database system (FDBS) architecture (see Sheth and Larson, 1990, for a description of the reference architecture) expressed using a client-server model. Below, is a brief description of the LII architecture followed by how it is applied in the system design.

### The LII schema hierarchy

The LII system architecture is founded on a multi-level schema hierarchy which is an extension of the standard ANSI/SPARC three-schema architecture (see Figure 3).

The local schema is, in effect, the data dictionary of the participating source database. Its data model is that of the database management system (DBMS) (e.g. if the source database is managed using Oracle, then the data model is relational).

The export schema is a representation of the local schema in a Common Data Model (CDM) used by the LII. The CDM will not, in general, be the same as that of a source database. The export schema describes only those parts of the source data that will be visible to the LII. Thus, confidential data in a source database can be obscured from access by the LII simply by not describing the data in the export schema.

The conforming schema is a representation of the export schema, again in the Common Data Model, but in a form that is semantically consistent with a globally conforming view of all source data described in the LII. This means, for instance, that if a property tax database and a Crown tenure database both contain data types called LandParcel, but they are not identical data types, at the conforming level they must be differentiated (e.g. perhaps into TaxLandParcel and CrownParcel).

Federated schemas can be thought of as virtual databases that represent data from more than one

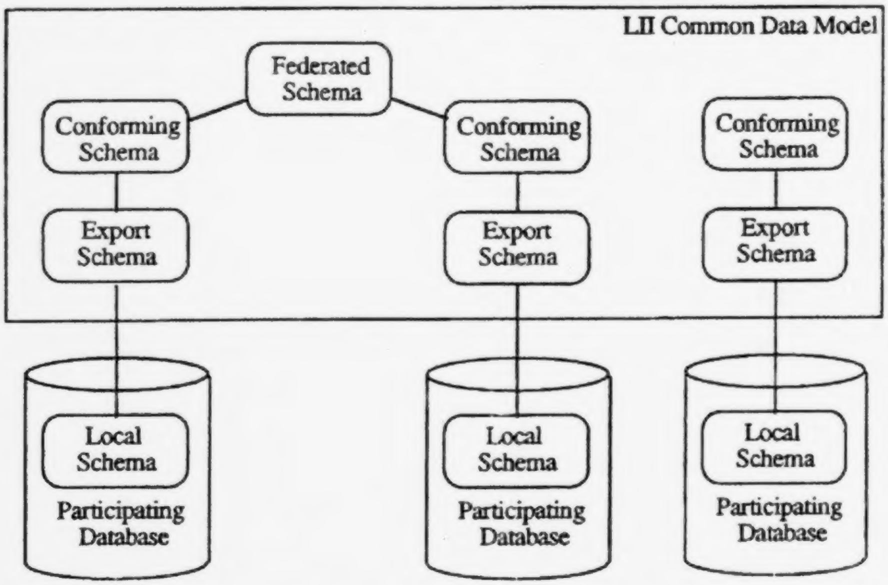


Figure 3. Multi-level Schema Hierarchy

physical source database. In other words, they are integrated views of more than one conforming schema. A user of the LII viewing a federated schema and ordering data based on its descriptions need not be aware that the data are being retrieved from different sources. So, for instance, a cadastral federated schema might consist of Crown cadastral data, private land title data, forest tenures and mineral tenures, each set of information being stored and managed in separate databases using different kinds of DBMS software and located in different places.

Although not shown in Figure 3, a complete schema hierarchy would, in general, show external schemas at the highest level. These are just like those defined in the ANSI model: different views of data maintained for different groups of users according to their needs. Typically, there would be one or more external schemas for each conforming schema and federated schema. The current design for the LII takes a simplified approach and disallows multiple user views, and so external schemas are collapsed into the conforming and federated schemas.

**The LII components**

Figure 4 shows the main components of the LII. Users interact with the LII via User Access Terminals (UATs) which act as clients to a number of servers. A UAT is a suite of software that can run on a variety of platforms. In addition to providing a human user interface, the

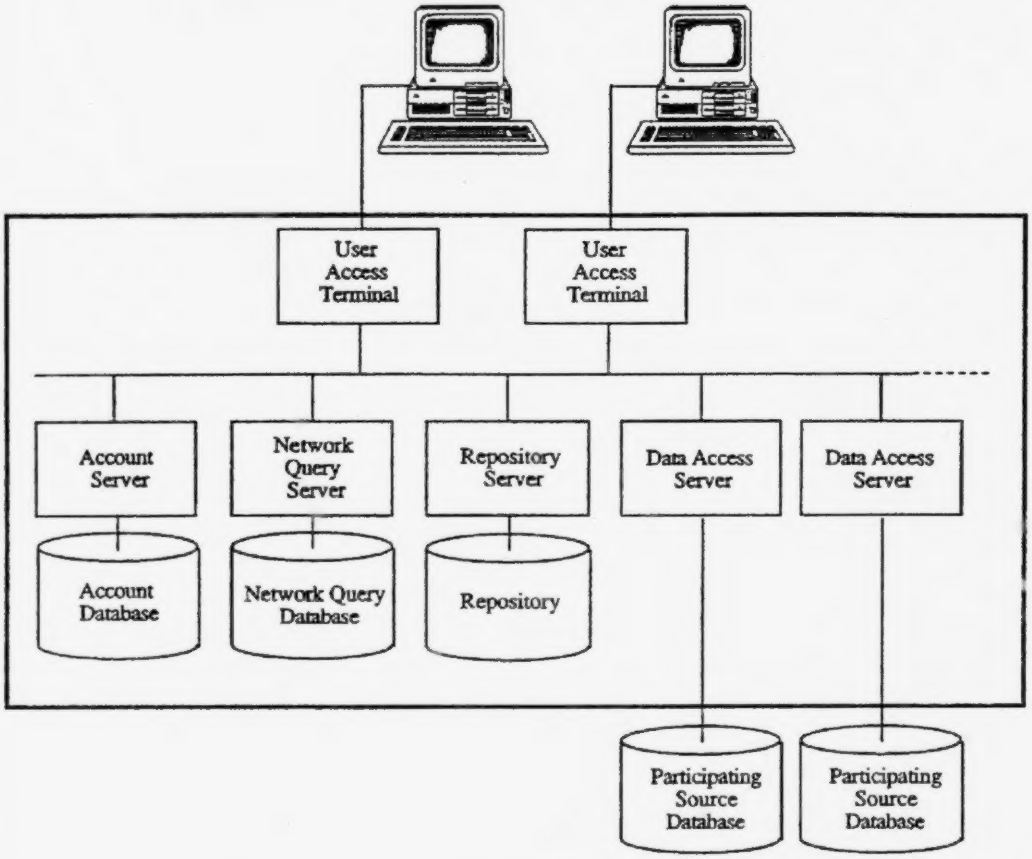

User Access Terminal
User Access Terminal
Account Server
Network Query Server
Repository Server
Data Access Server
Data Access Server
Account Database
Network Query Database
Repository
Participating Source Database
Participating Source Database

**Figure 4. The Client-Server Design of the LII**

UAT software may be invoked directly by application programs. They do this by calling services provided in a linkable object library of UAT services.

An LII user must hold an account on the system. These are much like user accounts managed by an operating system. In the case of the LII, accounts are maintained by the Account Server and its associated database. An account is a set of information about the user, including access privileges and quotas to system resources such as the processor and disk space, current levels of expenditure, and preferences for style of use.

An important component of the LII is the Repository and its server. This is, in effect, a global database of meta-data which describes at a number of levels the land-related data contained in the participating source databases and how the data can be retrieved. The Repository contains both textual descriptions of data, which the user may browse to learn about what land-related information is available (e.g. its subject areas, currency, geographical extent, custodial group) and also schema descriptions (e.g. data types and associated services, attributes and their domains), which enable the user to formulate queries for source data and enable the LII to process these queries. The former information is contained in the LII Directory and the latter information is contained in the LII Dictionary. The schema descriptions are those of conforming and federated schemas; these being at the top of the schema hierarchy and so the only ones visible to the LII user.

The Network Query Processor (NQS) handles the validation, manipulation and routing of queries from the UAT client, as well as processing land-related data retrieved from source databases and forwarding it to the Account Server for delivery to the UAT client. Queries are expressed in a Common Query Language (CQL). The database associated with the NQS contains export schema descriptions and provides temporary working storage for retrieved source data. The LII user does not have access to this database.

The Data Access Servers (DASs) are the interfaces between the rest of the LII and participating source databases. There is a DAS for every database connected to the LII. It handles syntax conversions between queries in the LII CQL that it receives from the NQS and the query language of the source database (e.g. SQL or a proprietary language of a particular DBMS or GIS). It also receives retrieved data from the database, repackages it if necessary, and routes it to the NQS for further processing. The cost of development of a DAS depends on the kind of source database to which it is connected and the level of data access required. Take for example, a DAS connected to a digital map database. If the DAS is required only to retrieve whole mapsheets, it will be simpler than if it were required to retrieve individual features within mapsheets.

## How the LII design meets key requirements

The LII design meets the key system requirements identified above in the following ways:

*Extensibility:* New land-related databases may be connected to the LII by adding new DASs. The Repository must also be updated to describe the new land-related information available for access. In principle, no other components need be changed.

The LII is also able to provide additional services for access to existing data sources. Assuming that the source DBMS is able to support the required level of access, a new service can be implemented simply by upgrading the attached DAS and updating the Repository accordingly.

*Flexibility:* At one level, the LII provides flexibility by shielding users from the physical and structural aspects of source data storage. This means, for instance, that if a government branch with a database connected to the LII moves to a different location, upgrades its computing facilities, or switches from a hierarchical DBMS to a relational DBMS, LII users will not need to know anything about the changes. (The Repository and/or the particular DAS may have to be updated, but this is not the concern of users of the LII.) One would hope that the only changes LII users see would be better response time and perhaps additional services for accessing this data set.

At another level, the LII provides flexibility by allowing the creation of custom data federations that answer the needs of specific user communities. This is done by creating federated schemas that provide integrated access to a number of data sources. As mentioned above, these simply appear as single databases to users. Federated schemas can be tailored to specific application areas, describing only the types of data required. As issues emerge and evolve, the federated schemas can be created and modified accordingly.

*Expandability:* The conceptual design for the LII outlined above may be translated into a variety of physical designs according to the size of the LII's user base, the number of source databases connected, the physical locations of users and databases, and the level of performance required. The first implementation of the LII will be modest, so a number of its components will reside on a single central host. However, as user demand develops and as more source databases are connected to the LII, it is likely that many of the LII functions will be distributed. Initially, this may mean expanding the host to provide symmetric multiprocessing. In time, better performance and reliability can be provided by distributing the server functions across many physical platforms, using perhaps some of the concepts of distributed DBMSs such a data replication management and data partitioning. In the longer term, the expandability of the LII may be limited by the communications infrastructure of the Province. Because it will manage the transfer of large amounts of data, the LII will need to use communications resources efficiently, and as the system expands, its physical design will reflect this need.

## Conclusion

The LII will provide the necessary services and information to enable access and facilitate the exchange and sharing of land-related data from many disparate sources in a heterogenous multi-database environment. It will provide useful services to users in the short term and will evolve to play a major role by providing the means of exchanging timely, specific and consistent land-related information within and between ministry enterprise systems.

The federated heterogeneous database system architecture supported by a client-server technology architecture provides for the extensibility, flexibility and expendability requirements for the LII necessary for the long-term viability of the system. It also ensures the autonomy of individual ministries in the management of their own databases.

The development of this physical infrastructure will require an understanding of both the organizations involved in the information sharing as well as the nature of the information shared. In addition, the sharing of information is only possible if there is a common semantic framework which encompasses all the ministries and Crown corporations of the Government. Sharing is a cooperative enterprise where success can be achieved only through the support and continued cooperation and participation of the agencies involved.

## References

BC Environment, Lands and Parks (1991): **Land Information Management Framework**, Volumes 1, 2 and 3, Victoria, B.C.

DMR Group Inc. (1989) for the B.C. Ministry of Crown Lands: **A Corporate Land Information Strategic Plan for the Government of British Columbia,** Volumes 1 and 2, Victoria, B.C.

Hofmeyr, H.D., D.E. Allen and E. Kenk (1991): **Setting the Stage for Data Sharing In British Columbia,** GIS'91 Symposium Proceedings, FRDA Report 153, Forestry Canada, pp. 89-93.

MacDonald Dettwiler (1991) for the B.C. Ministry of Crown Lands: **Land Information Infrastructure Statement of Requirements (High Level)**, Victoria, B.C.

MacDonald Dettwiler (1991) for the B.C. Ministry of Crown Lands: **Land Information Infrastructure Operational Concept**, Victoria, B.C.

MacDonald Dettwiler (1991) for the B.C. Ministry of Lands and Parks: **Land Information Infrastructure System Architecture**, Victoria, B.C.

MacDonald Dettwiler (1991) for the B.C. Ministry of Lands and Parks: **Land Information Infrastructure Target System and Migration Strategy,** Victoria, B.C.

MacDonald Dettwiler (1991) for the BC Environment, Lands and Parks: **Land Information Infrastructure High Level Design,** Victoria, B.C.

Sheth A.P. and J.A. Larson (1990): **Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases,** ACM Computing Surveys, September Vol. 22, No. 3, pp.183-236.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Efficient land information use within the province of British Columbia: Policy development based on need.

**Elaine M.J. Dawson**
**Robert Menes**
(presenting author)

Government Land Information Data Exchange
Surveys and Resource Mapping Branch
Ministry of Environment, Lands and Parks
1802 Douglas Street Victoria, B.C. V8V 1X4
Tel: 604-387-9353 Fax: 604-356-7831

## Abstract

*British Columbia is developing a Land Information Infrastructure system that will facilitate the sharing and exchange of land-related data within the Province. The sharing and exchange of land-related data is vital to support integrated resource management and the resolution of referrals within the Province. In order to manage and maintain the infrastructure and to ensure the consistent development of land-related information systems, the Ministry of Environment, Lands and Parks has developed operational policy, procedures, guidelines and standards for land-related information and technology management.*

*To determine what policy was required, the actual use of land information in government at an operational level was examined. From this study, what was needed at a management and policy level was ascertained. In order to develop policy that would work in a multifaceted government environment, it was necessary to: identify vendor-independent standards, permit varying levels of adherence, recognize the need for new organizational structures to mediate conflict resolution, accept the policy development as evolutionary and build upon existing government policy. The resultant land-related information and technology policies will enable the Ministry of Environment, Lands and Parks to fulfill its responsibility for ensuring the coordinated development, implementation, and approval of land-related information systems of government-wide interest and to develop an effective land-related data access and management system.*

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Design and operation of a Meta-GIS

**Ric Vrana**

Department of Geography and
Center for Urban Studies,
Portland State University
Portland, OR 97202
Tel: 503 725-5182
Fax: 503 725-5199

## Abstract

*Constructing spatial databases for resource management and mapping can be costly and time consuming. Systems integration, technology transfer, and efficient data collection can each be enhanced when private and public agencies cooperate in exchanging spatial data. The Northwest Land Information System Network (NWLISN), a consortium of federal and state agencies in Washington and Oregon, has contracted with researchers at Portland State University to build a Spatial Data Index (SDX). As a "meta-GIS" the SDX catalogs spatial data holdings by both subject theme and spatial extent. Queries and reports utilize both graphical and attribute descriptions of the data holdings of contributing members. Technical metadata issues addressed by this project include: 1) determining a suitable taxonomy for classifying data sets to facilitate description and querying, 2) graphic and topological approaches to the cross referencing of agency data tiled in differing geographic formats, and 3) tracking lineage, data quality, and currency of data sets. Institutional and organizational issues include: 1) coordinating funding and data reporting among multiple agencies, 2) determining a long-term strategy for the institutionalization of the index, and 3) implementing standards for meta data reporting.*

## Introduction

A catalog and indexing tool to help locate spatial data has been developed at Portland State University. This Spatial Data indeX (SDX), supports graphical and tabular queries to locate GIS and digital mapping data for a given areal extent and thematic subject. It is in effect, a meta-data GIS application, since it records descriptions of data sets which reside elsewhere.

Virtually all approaches to cost justification for GIS and LIS implementation recognize the characteristically high start-up costs associated with database development. (Dickinson and Calkins 1988, Wilcox 1990) Many applications require geodetic referencing and/or digital map conversion that can take months or even years to fully implement. (Epstein and Duchesneau 1984, Parent et al. 1989) Similarly, land records modernization and the building of large attribute databases for land information systems also is costly and time consuming. Reduced uncertainty in management decision-making, better planning data, and automated facilities management, are a few of the benefits of digital geographic databases. Because these are somewhat generic benefits, many agencies with interests in the same geographic region could potentially benefit from databases originally developed in another organization.

Despite the potential for sharing the benefits of commonly developed resources, a number of organizational and institutional impediments exist. (Dueker 1987) In particular, governmental organizations may replicate each other's spatial databases because organizational mandates and agency priorities differ even when areal extent and thematic subject matter are similar. (Chrisman and Niemann 1985) This results in many CAD, GIS, and LIS projects developing common base map and other data layers which have already been digitized or acquired by another agency. Therefore, as the technology matures, similar geographic databases proliferate throughout many levels and agencies of government within a region.

In the US pacific northwest, Oregon and Washington state agencies as well as federal agencies with regional headquarters in Portland or Seattle frequently require access to digital spatial data for mapping and geographic analysis. Because these data are expensive to generate, and duplicate copies of common themes are accumulating, it is increasingly important that digital map and attribute data be shared and/or traded whenever possible. The Northwest Land Information Systems Network (NWLISN) was established in 1988 as an ad hoc consortium of federal and state agencies to facilitate this exchange and inventory the state of geographic data processing in the Pacific Northwest.

Acting through the Oregon Department of Energy, a member agency, NWLISN has contracted with Portland State University to design, test, and initially to operate a database system that can assist its members in locating appropriate data sets from governmental agencies throughout the region. Such a system has several objectives.

- Avoid costly duplication of previously digitized map data.
- Promote inter-agency cooperation on data acquisition projects.
- Provide information on "most compatible" source products.
- Facilitate data standardization efforts and data transfer procedures.

By referring first to this index system, users identify data sets from other organizations that may be available for sale or trade in lieu of digitizing a source document. Planned data acquisitions are also cataloged to facilitate cooperation in obtaining and processing new digital data for areas of common interest. When more than one digital source is located using the index, additional attributes allow users to select the data which are most compatible with their system. Standardized methods of data description and reporting will assist the regional user community with adopting uniform data definitions and transfer methods based on the SDTS model. (USGS 1990)

Since the nature of the data is geographic, the index has been implemented with a GIS. In this way, SDX can be queried spatially or by attribute. When a geographic region of interest is specified, data holdings for that area identified by member agency and thematic content. Conversely, data themes such as land use, or soils, can be specified by query and all areas for which these data exist can be identified. In most instances, a query will consist of both location and thematic criteria, for example, when land use/ land cover data are requested for an area defined by a list of USGS 7.5' quadrangles.

This paper summarizes the essential characteristics of the design and operation of the NWLISN Spatial Data Index, herein referred to as SDX. Following this discussion, several more detailed points are offered concerning insights this project has provided for the technical and organizational management of spatial metadata.

## Basic architecture of SDX

The SDX can be thought of as three separate databases; the Agency Database, the Thematic Holdings Database, and the Spatial Cross-Reference Database. All are implemented in Arc/Info rev. 5.0.1 on a SunSparc 1 workstation at the Portland State University Urban and Public Affairs computer applications lab. The Agency and Thematic Holdings databases are also backed up as .DBF compatible files on a Macintosh PC running FoxBase 2.0. Copies of these relational files have been sent to the Washington State Library in Olympia to support some tabular queries and reporting from that location.

### Agency database

The Agency Database provides information on participating agencies, their location and designated technical contact person. The agency contact is someone who has been identified as responsible for reporting data holdings to the SDX and is available to handle follow-up inquiries about data availability from interested parties.

In addition, all GIS and mapping systems in use at an agency are listed here, together with information concerning their computing platforms and operating systems. Preferred export formats and supported magnetic media are listed in the index so that persons contacting these agencies know what logical and physical formats to request.

Agencies with data holdings reported to SDX consist primarily of NWLISN members. In a few cases, non NWLISN member agencies affiliated with the Washington Geographic Information Council have been incorporated as well. For the most part all these agencies are either federal/regional or state-wide in scope but at least one county and another two-county regional planning agency is currently represented. Not all agencies contained in the Agency Database list data holdings with SDX. Some agencies are participating with the expectation that future database building efforts might be cooperatively arranged with others seeking similar data.

Figure 1 provides a summary of agencies listed with SDX and the GIS or CAD systems in use. Note the particular dominance of one vendor's GIS product among agencies in the Pacific Northwest. This is not as monolithic as it seems, however, when it is remembered that data are still structured and referenced quite differently from one installation to another. In addition, there are enough other systems in use to require that SDX incorporate non-vendor-specific descriptions for data holdings and extents that could be readily understood by this heterogeneous user community.

**SDX *Agency Database***

| | | |
|---|---|---|
| | 12 | Federal |
| | 7 | Oregon |
| | 9 | Washington |
| Total Agencies = 29 | 1 | Other |
| Total Systems = 47 | 25 | Arc/Info |
| | 9 | Intergraph |
| | 4 | MOSS |
| | 2 | GRASS |
| | 1 | AutoCad |
| | 1 | Synercom |
| | 1 | I2S |
| | 1 | MapInfo |
| | 1 | EPPL7 |
| | 2 | In-house... |

**Figure 1.**

SDX agencies and spatial data systems in use

### Thematic holdings database

The Thematic Holdings Database consists of records describing individual data sets reported by participating agencies. Data sets are uniquely identified by a combination of owning agency, the data set name, and completion status. (Figure 2) Additional attributes of each data set include:

- Thematic subject code (see below)
- Up to three keyword identifiers
- Tiling method
- Compilation data (when data were originally made digital)
- Compilation source or provider
- Compilation scale
- Revision date
- Host system.

**SDX *Thematic holdings database***

**Identified by name, agency, status**

**Attributes:**

**Subject coded by theme**
**Additional key words for search**
**Compilation/Revision dates**
**Source map or base data**

**Status Codes:**

| | |
|---|---|
| C | Completed |
| I | In Progress |
| P | Planned (Budgeted) |
| D | Desired (Wish-list) |

**Figure 2.**

Thematic holdings database: principal attributes and status codes

The status field is an important one for two reasons. A high priority of NWLISN is the ability to utilize the system to identify planned data acquisitions in other agencies, in order to find partners for shared co-development. Also, by using this field as part of the unique identifier, an agency can list two data sets with identical names, one complete and one planned. Four valid codes for the status field exist, corresponding to" Complete", or "In progress", and two categories of future acquisitions, "Planned" in the sense of budgeted in the current cycle, and "Desired" for those acquisitions on the planning horizon but not yet budgeted or prioritized. Agencies therefore can get a short and long term view of future data developments in the region.

The Thematic Holdings Database is a relational file and is linked to the Agency Database through the agency acronym fields of the two databases. When used alone, these two databases can provide information about the existence of individual data sets and who to contact to learn more about them. What is not contained in these listings is a description of the areal extent covered by individual data sets.

### Spatial Cross-reference database

The extent of an individual data set is maintained in the Spatial Cross-Reference portion of the SDX. Contributing agencies manage geographic data in a variety of tiling

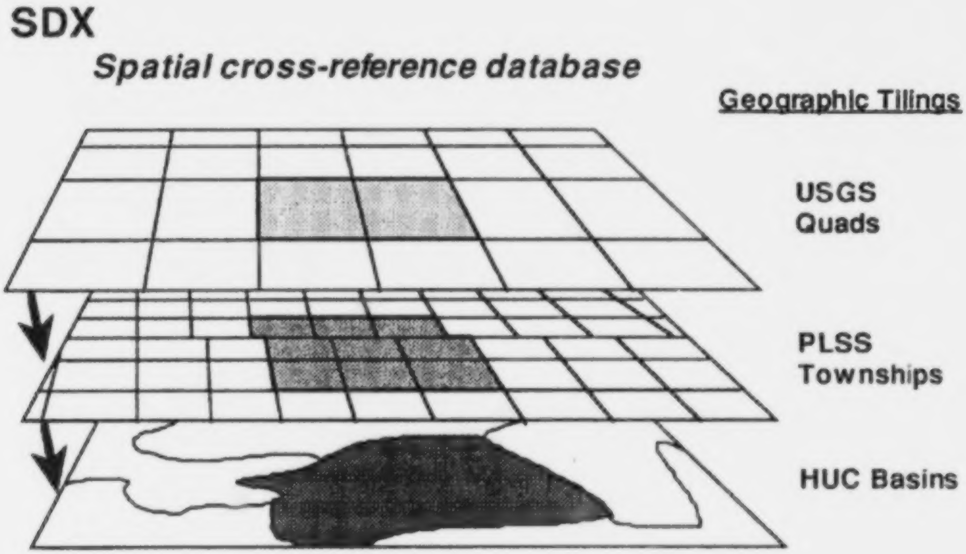


**Figure 3.** The function of the spatial cross-reference database

schemes. USGS quadrangles, Public Land Survey System (PLSS) townships, and Hydrologic Unit Basins are some of the more prevalent views of geography used by NWLISN agencies. When a data set is reported to SDX, the tiling system used to reference it is contained as an attribute. For each listed data set, areal extent is provided as a list of these quadrangle, township, or basin units. In cases where one of these three tiling systems are not employed, an approximation, usually in terms of 7.5' quadrangles, is used instead.

The function of the Spatial Cross-Reference Database is to find data sets meeting the thematic and geographic requirements of a user even if these data are cataloged in another tiling system. Given a list of quads, for example, the index can locate data for that area which are available from an agency reporting its holdings only in terms of townships. (see Figure 3)

Some implementation details of this cross-reference system are addressed in the technical issues section of this paper. It is a problem that is characteristic of the SDX user community, which is composed of agencies with widely differing views of the same geographic space. Users locating the data meeting their needs can contact the appropriate agency and request data in terms familiar to the holder. This facility enables participating agencies to report data holdings in accordance with the tiling schemes used in-house. Extent reporting thus remains in user-native terms when possible, and for GIS installations where a data set is referenced to a set of tiles, it can be accomplished by simply obtaining a list from the database.

### Operation of SDX

Design and initial data collection for SDX has been completed, although additional query and reporting procedures are still being developed, particularly where frequent transactions can be automated. In its first phase, the SDX is a single site application, residing at Portland State University. Inquiries are handled via telephone or written communication and tabular and/or graphic results are provided within a short turn-around period. In addition to the query service, quarterly reports of the index contents are issued to subscribing agencies.

Query service access to SDX is not limited to NWLISN members at this time. There has been some discussion of a fee for access by non NWLISN agencies after the initial trial phase. Further details for the future of SDX are addressed in the institutional issues section, below.

When someone who needs information about geographic data calls the service, a staff person translates the request into a query to SDX. This is done as a three step process consisting of a specification of the geographical extent to be searched, the subject coding or other thematic attribute to be matched, and the type of reporting format desired by the user. These are summarized in figure Four.

*Querying* **SDX**

**Step 1.** Specify geographical extent to be searched
- A. Oregon, Washington, both
- B. List of quads, townships, basins
- C. County
- D General Geographic area (approximate w/ quad list)

**Step 2.** Specify thematic attributes for the search
- A. Subject code
- B. Status (Complete, in-progress, Planned, Desired)
- C. Agency owning the data
- D. Tiling unit
- E. Scale
- F. Combination of above &/or date, source

**Step 3.** Specify desired reporting format
- A. Report on data set Information (by agency, status, code, scale, etc.)
- B. Listing of geog. extent by tiling units
- C. Plot of geographic extent of data set
- D. Print-out of related Agency Information

**Figure 4.**

Steps in querying the spatial data index

## Technical metadata issues for spatial index systems

SDX is a GIS application that handles descriptions of GIS and cartographic databases residing at separate sites. A broad view of metadata defines the concept as data that describe other databases. (Elmasri and Navathe 1989, p.489) Most metadata management systems are in the form of data dictionaries and catalog applications that store structural and semantic information about individual databases as schema descriptions and object definitions. Comprehensive information integration between two databases requires a detailed look at these metadata in order for the meaning and structure of data to be preserved. (Nyerges 1989) SDX is not such a tool. Rather, it can be viewed as a region-wide metadata tool, providing the first step toward better data integration by identifying relevant aspects of data sets for consideration.

What then, are some of the metadata constructs embodied in the SDX? They involve definitions and structural constructs that facilitate thematic, locational extent, and data quality description of whole data sets.

### Taxonomy for data set subject classification

Developing a standardized approach to describing the thematic content of GIS and cartographic data is not a terribly straightforward task. Descriptions of data tend to be couched in terms familiar to specific software user communities and will vary considerably among agencies with distinctly different mandates. A coverage to an ARC/INFO user may hold quite different thematic data than a design file to an Intergraph user.

A typology for thematic data layers was developed after examining a number of data cataloging efforts in several other states. Although no one uniform scheme has been widely adopted, the two basic approaches describing thematic data are the use of subject codes and keywords.

Subject codes have the advantage of being hierarchically ordered and can support range specified queries, thus adapting to the level of specificity requested by a user. As with any hierarchical coding scheme, (SIC, SLUC etc.) inevitably there will be cases that are difficult to classify. The intent is to be complete and exhaustive.

Every listed data set must be identified with a single subject code. Subject coding of data sets is done by the reporting agencies according to a code sheet developed by SDX staff. Three digit integer codes depict seven main divisions of thematic data: Base Reference, Natural Features, Cultural Features, Socio-Economic, plus Areal, Satellite, and other Imagery. Each subject code within a main division, such as hydrology within natural Features, or State and County Boundaries within Cultural Features, begins with the same digit. Each division may be further broken out into broad classes ending in zero and more specific subsets of those classes, (Slope = 215 of Topography = 210, of Natural Features, 2 - -) Synonyms, explanations, and comments accompany each class to help users adopt a fairly uniform approach to categorizing their data. It is recognized that this is an imperfect method, but one which retains strong user participation.

Keywords are additional descriptive 'tags' for each data set. Keywords are not hierarchically ordered and so present difficulties for automating database searches for general to specific ranges, but they have the advantage of being intuitive and mnemonic. The SDX provides up to three keyword identifiers for each data set. In a sense they are a compromise, and can be used either to further specify the subject code (225 wetlands + "Estuarine") or data source information (213 DEM + "DMA" and "250K") or important related tabular items or files associated with the spatial coverage, (328 Census Tracts +"Income" and "Education").

## Extent description and cross-referencing spatial tiling schemes

Several possibilities for metadata describing data set extent exist. Spatial index systems in Santa Barbara, California and Columbus, Ohio employ polygonal descriptions to which thematic data attributes are tied. (RLG News 1989, Marble 1991) Catalogs in some other states are relational files only and do not feature a graphical description of the extent of data holdings, only listing counties or quadrangles for which data exist.

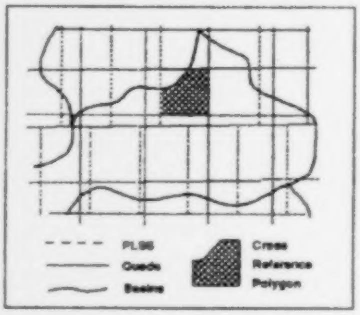
It was recognized at the outset of the NWLISN project that neither of these approaches would suffice. Polygonal boundaries are most useful for small project GIS applications and most NWLISN agency work involves very large areas in the two states or small areas that are widely scattered. (Much public landholding in alternate PLSS townships and/or sections). Building a relational database alone would forgo the graphical query and reporting that was desired, as well as make cross-referencing between tiling schemes very cumbersome.

Tiling spatial data refers to the partitioning of extensive map coverages into files of convenient size for storage and processing. Automated mapping and GIS have adopted previous map sheet partitioning strategies to the digital tiling equivalents. The result is that much digital cartographic data have been converted from map sheets directly into digital tiles with accompanying positional and attribute mismatches along the edges. (Chirsman 1990) Nevertheless, even as computing platforms and software advance in capability and speed, there are institutional and practical reasons why most geographic information professionals continue to think of geography in terms of the tilings appropriate to their agency mandates. Virtually all USGS Mapping Division products are can be described in terms of the basic 7.5' quadrangle building block. Meanwhile, hydrologic databases at the Water Resources division of USGS is stored in hydro basin files from the state HUC map series. Much Census Bureau data, while tiled by counties, can be referenced to the original quadrangle DLGs used to create the 1990 TIGER files. Other state and federal agencies use the Public Land Survey system townships as a tiling method.

Forcing all data set extent descriptions to conform to a single standard such as quadrangles would create an extra burden on some agencies to translate only for the purpose of contributing a listing. SDX users finding those data potentially useful would then call back, requesting data sets in terms of a non-native format, causing another round of work.

The solution to this problem was to use GIS overlay processing to enable data encoded in one spatial tiling form be cross referenced to data coded in another. (Figure 3) A practical limit was imposed, however, because if a large number of tiling schemes were employed, or if irregular polygonal boundary descriptions accepted, the integrated cross-reference coverage layer would contain and unmanageable number of reference polygons. Therefore, the three most prevalent tiling methods in use among NWLISN members were chosen. These are the USGS 7.5' quadrangles, PLSS townships, and HUC basins. Coverages defining these units are intersected into an integrated cross-reference layer which contains over 13 000 reference polygons for the state of Washington and over 10 000 for Oregon. (see Figure 5)

**Integrated layer of the cross-reference database**



- Cross Reference Polygons link to parent quad, township/range, hydro-basin
- Thematic data set IDs tied to individual quads, township/range, hydro-basin

**Figure 5.**

Cross-referencing various tiling schemes describing spatial data sets

An individual data set record from the Thematic database is related to a link list of many quads, townships, or basins. An individual reference polygon is linked only to its parent quad, township, and basin and not directly to all metadata records in the thematic database that may occupy that sliver of geographic space. The reference tile lists are then related to the link lists for the thematic metadata to accomplish the cross-referencing. (see Figure 6) This method accomplishes the cross-referencing with a minimum number of separate data records (still considerable !) and results in reasonable query response times. It does this at the expense of spatial precision, however. If, for example, a query for all data is expressed as a list of townships, whole basins with those data that happened to correspond with even a corner of one of the townships will be selected. In most cases, this discrepancy is not very severe. Since the method cautiously errs on the side of over-, rather than under-reporting, the added benefit of presenting the inquiring user with query results which directly reference the native tiling scheme of the originating database is desirable.

## Metadata describing data quality

A third major technical issue to be addressed in a spatial metadata indexing system is how to catalog and report data quality and accuracy information. In this regard, SDX has not implemented a satisfactory "solution". In part this is a consequence of the nascent stage of standardization in reporting data quality. Furthermore, since the index is operated by a third party without direct access to the data, there is no independent way to verify statements about the quality of data sets listed by member agencies.

A standard for conveying useful information concerning data quality most likely will emerge from efforts to implement the national Spatial Data Transfer Standards (SDTS) (USGS 1990) The standard specifies that statements accompany file transfers concerning five components that document data quality.

- The lineage of the data; source documentation and processing.
- The integrity of positional accuracy in the data.
- The accuracy of attribute classifications in the data.
- The logical consistency of the data; how well the data conform to the model schema.
- The completeness of the data as described in accompanying documentation.

While the standard has undergone many iterations of review and comment there does not yet exist a practical software tool to implement SDTS. SDX's *caveat emptor* approach to data quality therefore, reflects the state of the discipline with regard to this issue. But a "fitness for use criteria" can be employed only if metadata contain enough attributes to form an expectation of the integrity of the data. Certain fields in the SDX Thematic Holdings database can be evaluated in just such a way. Source document, compilation scale (of the smallest scale component that went into the digitizing), and compilation and revision dates, are all attributes that can be queries for any data set description and reviewed by prospective users.

## Institutional metadata issues for spatial index systems

The technical considerations discussed above, for the most part describe how SDX works as a GIS metadata handling application. How SDX functions in the social and political milieu of spatial information sharing raises several interesting institutional metadata issues for consideration. Three of these are; how to coordinate the many individual agencies contributing to SDX, the closely related issue of how SDX is to evolve in the future, and the wider issues of the standardization of spatial metadata.

### Coordination among multiple agencies

Informal arrangement for data sharing have persisted among users since the dawn of digital cartography, but the scale and scope of the current activity far exceed the ability of informal constructs and personal contacts to rationalize it. Nolan and Nyerges (1991) note that this cooperation is successful only to a point, beyond which, some coordination is required in the form of more formal efforts. These researchers proposed a Land Information System Network to establish organizational and technical linkages between local and regional agencies with interests in managing land records and geographic data. Such a network would not only exchange digital data but allocate responsibility for maintaining layers among participating entities.

The NWLISN is a less ambitions undertaking, relying on the voluntary cooperation and good-will of interested parties within participating agencies. The consortiumwas established through a memorandum of understanding where each agency sends a representative to a Steering Committee which establishes policy. Technical working groups implement these policies; one of which is SDX. PSU was adopted as a contractor to design and initially operate SDX, and a NWLISN Index subcommittee meets periodically to oversee the venture. Thus, contacts among NWLISN member agencies and between them and PSU occur at a variety of levels. Managers and administrators at the Steering Committee level are enthusiastic about the index due to government mandates to cooperate on data acquisition. Representatives at the technical and index subcommittee levels are also enthusiastic but realize that the project involves a good bit of work. Each agency names a contact person to handle inquiries and data reporting. For the first year of the project, reporting has been slow since the project is lower on the horizon of people who have not been involved in planning and design meetings and are also working on day to day production.

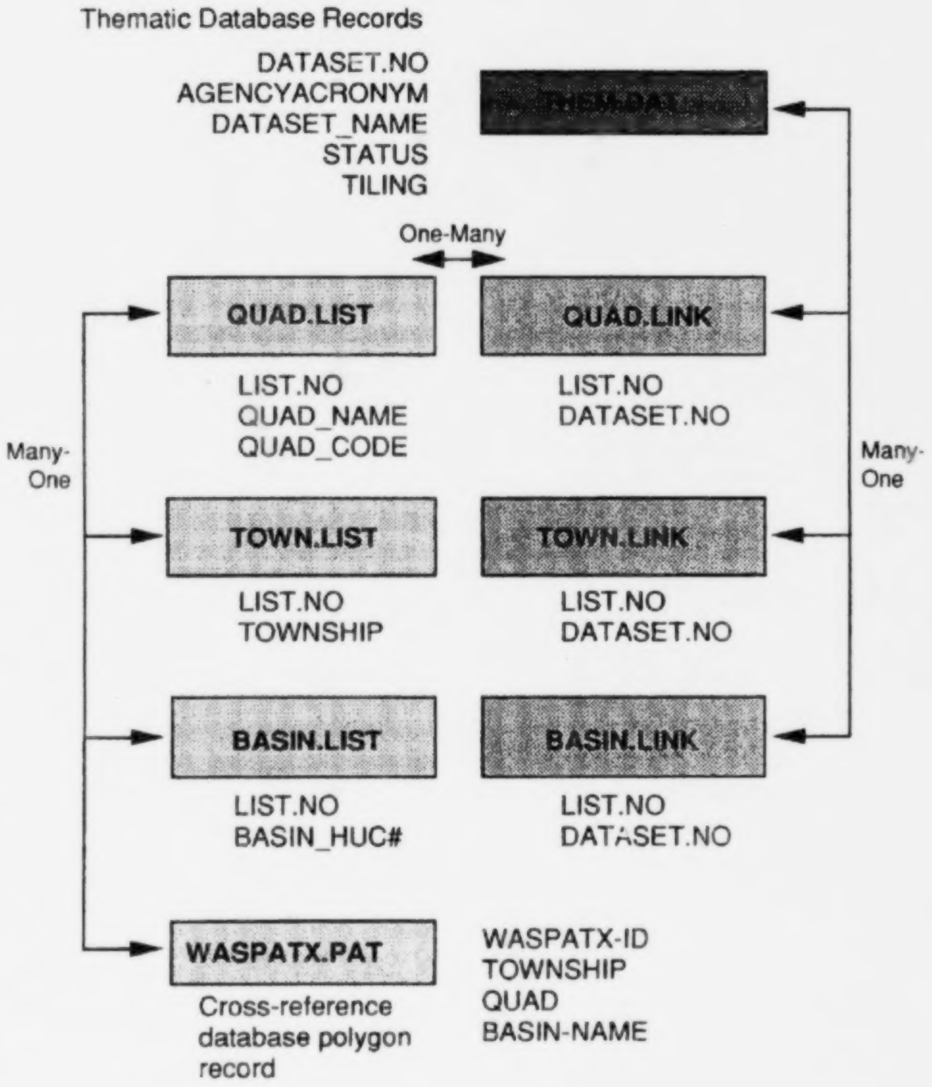
Managing the SDX therefore has required some vigilance to keep participating members motivated and informed of progress. That this has come as far as it has in the absence of a strong central unifying organization is a tribute to the recognition that a regional metadata resource on spatial data will be beneficial to many.

### Planning for long-term deployment of SDX

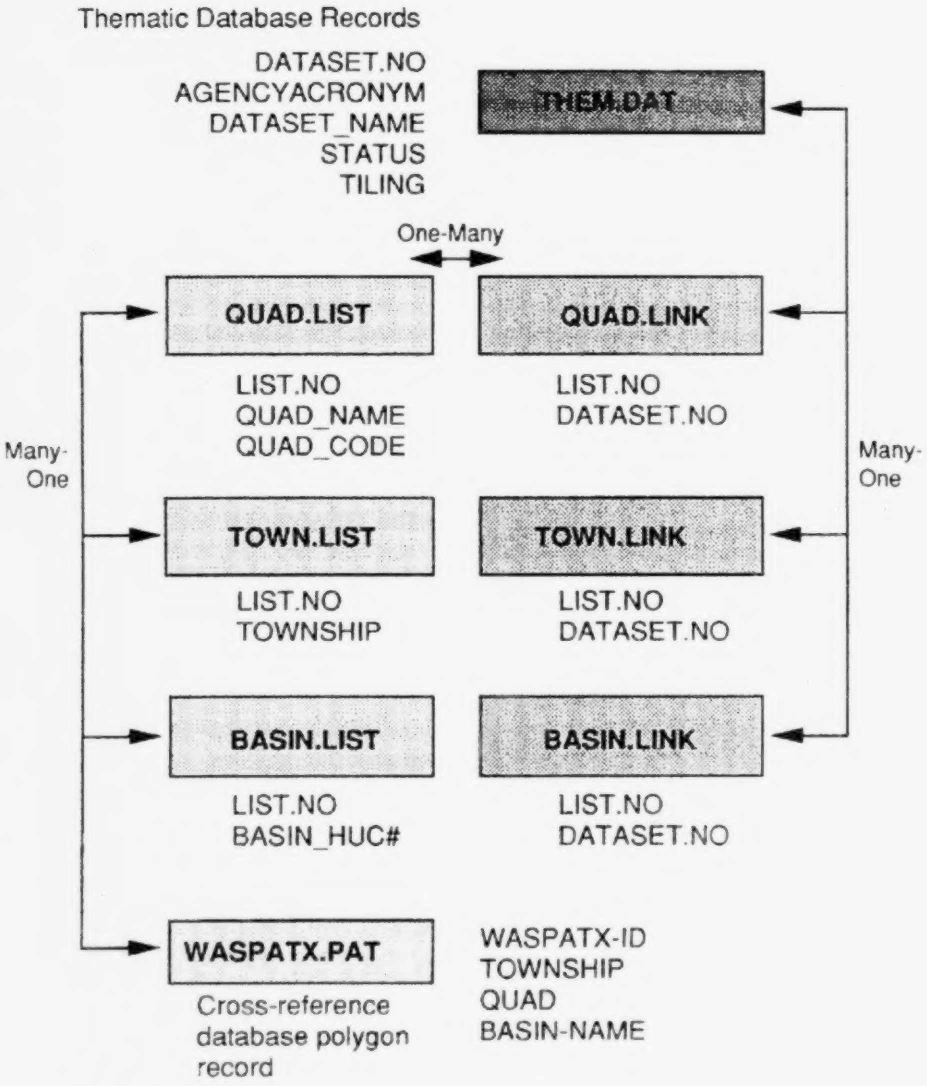
After the initial development of SDX, what organizational structure can maintain and operate it in the long run?

Currently the Washington and Oregon portions of SDX reside as separate and identical applications at PSU. The NWLISN has decided to seek a more permanent home for the application in an Oregon and a Washington State institution. SDX staff and the index subcommittee have formulated a redeployment plan based on incremental steps using in-place technology to transfer the facility to the Oregon State Service Center for GIS in Salem, and the Washington State Library in Olympia. Researchers in Portland remain interested in further developments, NWLISN wishes to continue its

# Thematic to Spatial Cross-Reference Database Link Schema


Thematic Database Records
DATASET.NO
AGENCYACRONYM
DATASET_NAME
STATUS
TILING
THEM.DAT
One-Many
QUAD.LIST
LIST.NO
QUAD_NAME
QUAD_CODE
QUAD.LINK
LIST.NO
DATASET.NO
Many-One
Many-One
TOWN.LIST
LIST.NO
TOWNSHIP
TOWN.LINK
LIST.NO
DATASET.NO
BASIN.LIST
LIST.NO
BASIN_HUC#
BASIN.LINK
LIST.NO
DATASET.NO
WASPATX.PAT
Cross-reference database polygon record
WASPATX-ID
TOWNSHIP
QUAD
BASIN-NAME

## Thematic to Spatial Cross-Reference Database Link Schema


Thematic Database Records
DATASET.NO
AGENCYACRONYM
DATASET_NAME
STATUS
TILING
THEM.DAT
One-Many
QUAD.LIST
QUAD.LINK
LIST.NO
QUAD_NAME
QUAD_CODE
LIST.NO
DATASET.NO
Many-
One
Many-
One
TOWN.LIST
TOWN.LINK
LIST.NO
TOWNSHIP
LIST.NO
DATASET.NO
BASIN.LIST
BASIN.LINK
LIST.NO
BASIN_HUC#
LIST.NO
DATASET.NO
WASPATX.PAT
WASPATX-ID
TOWNSHIP
QUAD
BASIN-NAME
Cross-reference
database polygon
record

involvement to insure standardized access and reporting procedures for both states, and some redesign and coordination will still be necessary in the beginning phases. Therefore, redeployment will include transaction logging and evaluation in order to direct further development.

In Olympia, SDX will be initially a relational database application without graphic query capability or spatialtile cross-referencing. Evaluation of this service will demonstrate the need for more spatial processing and visualization. Depending on user demand, the index could then be supported on a GIS at the State Library, or elsewhere, with the Library acting as an interconnected node.

In Oregon, the State Service Center for GIS is also an ARC/INFO site and can already support SDX as is. This agency is interested in establishing a state-wide network of Arcview (browse only) nodes to support use of its many services, including a version of SDX.

Additional policy decisions concerning non- NWLISN access fees, subscriptions, and restrictions may be appropriate and will be made in accordance with open records laws in both states.

### Development and implementation of spatial metadata standards

The NWLISN Spatial Data Index is one implementation of a data cataloging system on a regional level. It has involved the participation and cooperation of many agencies which use different automated systems, data definitions, and geographic partitioning strategies. Methods of collecting, representing, and disseminating spatial metadata have been central to the technical and institutional decisions discussed above. As a more integrated culture of geographic information processing emerges, the need for metadata describing spatial databases will expand. Interest in the role and utility of metadata is rising in many disciplines. (Medyckyj-Scott et al. 1992) It is not too early to consider the requirements for a generic standard for spatial metadata databases and applications.

Momentum has been building in recent years for a standard approach to spatial data exchange. Metadata to describe cartographic and geographic databases is therefore a crucial part of the SDTS (Spatial Data Transfer Standard), currently in the process of becoming a federal information processing standard (FIPS) (McDermott 1991). The standard has contributed much to the definition of a language describing geographic and cartographic entities, and provides a means to address data quality issues. It prescribes formats for data transfer modules and the metadata to describe their contents.

Because the principal focus of SDTS is the actual data transfer process, another standards process has been initiated to complement SDTS with a comprehensive specification for spatial metadata. (ASTM 1991) This effort, which is being moved through the ASTM standards process, may evolve as a voluntary guideline rather than a national processing standard. Nevertheless, it has provided a convenient forum for the discussion of spatial metadata issues.

Discussions regarding this draft specification in part have clarified the scope of how spatial metadata is useful in geographic information processing. Several interests related to spatial metadata imply slightly different points of view regarding the use and development of a spatial metadata standard. These are:

- Metadata for cataloging and indexing spatial data from a variety of agencies.
- Metadata to accompany an individual data transfer (SDTS)
- Metadata to document internal processing and operations on data, and
- Vendors of GIS and other software who many implement parts of the standard.

The SDX project corresponds to the first of these four interests. The essential question to be addressed by this aspect of spatial meta-data is, "What constitutes a description of a spatial data set?" Vrana (1991) has proposed a series of records to describe aspects of a data set for indexing purposes. t contains ten categories of information with nested subcategories indicating information relating to the "who, what, where, and how" of spatial data. (See table 1) The author participated in discussions with a variety of individuals either engaged in data indexing tasks or involved with organizations, like NWLISN, which have a serious interest in developing metadata to facilitate data exchange. For the most part, these recommendations reflect these conversations and the experiences and lessons of the SDX project regarding spatial metadata.

**TABLE 1**

**Spatial Meta-Data Fields to Support Data Indexing and Catalogs**

1. Data set identification
    a. Each data set to be referenced by name or numeric ID
    b. Individual data sets may be uniquely identified by a compound key composed of the data set name, owner (agency), and possibly its availability status code (see below, 4a)
2. Thematic description of the data set
    a. Data theme represented by subject code — and/or
    b. Data theme represented by keyword descriptor

3. Earth location description of the data set
    a. Generalized text description. (i.e. "North east Oregon")
    b. Polygonal boundary of area for which data set exists
    c. Method of tiling or partitioning to store/ manage the data (e.g. 7.5' quads, Townships, etc.)
    d. Listing of all tile units covered fully or partially by the data set
4. Temporal description of the data set
    a. Availability status code (e.g.: C= completed, I= in progress, P= planned, D= desired.)
    b. Date of original compilation as a digital data set
    c. Date of latest revision
5. Source characteristic of the data set
    a. Source document type (i.e. map, air photo, other)
    b. Agency responsible for original compilation or survey
    c. Date of source document
6. Spatial framework of data set
    a. Spatial Model
    b. Scale of compilation (smallest scale from which data were derived/conflated)
    c. Coordinate system
    d. Projection
7. Quality characteristic of the data set
    a. Data lineage availability
    b. Positional data accuracy check
    c. Attribute data accuracy check
    d. Check for logical consistency
    e. Check for completeness
8. Media/product characteristic of the data set
    a. Program or software that created the data set
    b. CPU and operating system
    c. Format availability (if supplier has a choice, indicate preferred, or native format)
    d. Data set size
    e. Information on cost, licensing, restrictions etc.
9. Supplier/producer characteristic (Contacts for more info and/or exchange)
    a. Agency name and division
    b. Address/ routing
    c. Contact person and telephone.
10. Additional characteristics pertaining to image data sets

    These elements pertain to digital imagery *in addition* to the elements that apply for all spatial data.

    a. Type of digital imagery [I. = aerial photography; II = multispectral scanned imagery, radar, etc.]

**On I.** b. Scanned image, or digital rectification
    c. Scale
    d. Flying height (focal length)
    e. Film type (B&W, IR, Color, CIR, etc.)
    f. Film format (size)
    g. Sensor aspect (vertical, oblique, stereo, panoramic)
    h. Clarity (% cloud cover, dust, etc.)
    i. Date of exposure

**On II.** b. Platform (Landsat, Spot, others, etc. possibly airborne)
    c. Minimum resolution (pixel size in ground units)
    d. Flying height/ altitude
    c. Spectral band(s) (in wavelength units OR another common band reference scheme)
    d. Scene identifier (as referenced by source. Data set identifier(1.) might be more mnemonic)
    d. Scene size (frame reference as "earth-location" above. This is file size)

## Conclusion

The NWLISN Spatial Data Index project has developed an application that can reference a variety of data holdings from agencies with different mandates, thematic subjects, and geographic views. Designing the application and making it work among a heterogeneous user base has required an examination of a number of technical and institutional issues discussed in this paper.

SDX and other prototype data indexing systems can be considered partial models for the development of regional and larger metadata handling systems. To accomplish this, the lessons and experiences of projects such as SDX should be examined with the intent to establish generic and uniform methods for representing spatial metadata, perhaps leading to a specification of a standard.

The ten categories offered above are put forward as the basis for discussion among parties interested in the spatial metadata question. Increased access to the data resources of public domain and other agencies on a quid-pro-quo basis will be facilitated with a standardized approach to data set description such as this.

## Acknowledgements

The development of SDX came about through the efforts of many individuals connected with Portland State University or NWLISN member agencies. The author wishes to express his special gratitude and appreciation for the assistance and direction provided to the project by Cary Lorimor, Margo Blosser (especially for figure 6), Steve Sherer, and Dick Lycan. Funding for the project was obtained from member agencies and administered through the offices of the Oregon State Service Center for GIS, a unit of the State Department of Energy.

## References

ASTM D18.01.05 Subcommittee on Spatial Metadata Standardization. (1991). *Draft Proposed Specification for Meta-Data Support in Geographic Information Systems.* Presented at GIS/LIS'91, Atlanta, Oct. 29, 1991.

Chrisman, N. (1990). "Deficiencies of Sheets and Tiles: Building Sheetless Databases." *International Journal of Geographical Information Systems.* 4:2, 157-167.

Chrisman, N. and B. Niemann. (1985). "Alternative Routes to a Multipurpose Cadastre: Merging Institutional and Technical Reasoning". *Proceedings of Auto-Carto 7,* International Symposium for Computer-Assisted Cartography, American Congress on Surveying and Mapping, Washington DC, March 1985, pp. 84-94.

Dickinson, H.J. and H. Calkins. (1988) "Economic Evaluation of Implementing a GIS." *International Journal of Geographical Information Systems.* 2:4, 307-327.

Dueker, K. (1987). "Multipurpose Land Information Systems: Technical, Economic, and Institutional Issues." *Photogrammetric Engineering and Remote Sensing.* 53:10, 1361-1365.

Elmasri, R. and S. Navathe. (1989). *Fundamentals of Database Systems.* Redwood City, CA: Benjamin/Cummings Publishing.

Epstein, E. and T. Duchesneau. (1984). *The Use and Value of a Geodetic Reference System.* Research study by University of Maine at Orono. Available from National Geodetic Information Center, NOAA. Rockville, MD.

Marble, D. (1991). Personal communication.

McDermott, M. (1991). "The Spatial Data Transfer Standard." *Technical Papers of the ACSM-ASPRS Annual Convention,* Baltimore 1991. Vol. 2, Cartography and GIS/LIS. pp. 222-230.

Medyckyj-Scott, D.K., I. Newman, C. Ruggles, and D. Walker. eds. (1992). *Metadata in the Geosciences.* Loughborough, UK: Group D Publications Ltd.

Nolan, T., and T. Nyerges. (1991). "Land Information Systems Network in the Puget Sound Region". *URISA Journal.* 3:2, 25-34.

Nyerges, T. (1989). "Information Integration for Multipurpose Land Information Systems." *URISA Journal.* 1:1, 27-38.

Parent, P., Bruce Joffee, and Robert Finkle. (1989). "Estimating the Costs of Building your AM/FM Database." *GIS/LIS'89 Proceedings* Vol. 1, 143-151.

RLG Inc. (1989). "RLG Enters New Sphere with Geoinformation Project." unsigned article, *RLG News.* No. 19 Spring 1989. 1200 Villa St. Mountain View, California 94041.

USGS. (1990). *The Spatial Data Transfer Standard.* Draft standard, version 12/90. Available from USGS-National Mapping Division, SDTS Technical Review board.

Vrana, R. (1991). "Spatial Data Indexing Concerns for a Meta-data Standards Specification." Report submitted to ASTM D18.01.05 Subcommittee on Spatial Meta-data Standardization, GIS/LIS '91. Atlanta, Georgia, Oct. 29.

Wilcox, D.L. (1990) "Concerning 'The Economic Evaluation of Implementing a GIS'." *International Journal of Geographical Information Systems.* 4:2, 203-210.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# The use of GIS as a dynamic multiple-use management tool by the U.S. Bureau of Land Management on federal public lands in Colorado

**Bob Moore**

Colorado State Director
Bureau of Land Management
2850 Youngfield Street
Lakewood, Colorado 80215

## Abstract

*The United States Bureau of Land Management (BLM) currently administers more than 270 million acres of public lands under the management principles of multiple use and sustained yield, as mandated by the Federal Land Policy and Management Act of 1976 (FLPMA). A wide variety of natural resource management techniques have been used over the 45-year history of the agency and new methods are constantly being evaluated for their potential to improve overall efficiency and meet increasing demands.*

*In 1982, responding to the dramatic increase in energy development activities for coal and oil shale resources, the Colorado State Office of the Bureau of Land Management initiated the use of Geographic Information System (GIS) technologies to support the resource management planning process. As this state-of-the-art technology has evolved over the years, the BLM has continuously incorporated new automated procedures to assist in the development of management strategies and in the production of public documents containing detailed maps and statistical reports.*

*As a result of the successes and frustrations of GIS implementation in Colorado, the BLM is learning new ways of efficiently incorporating this technology into its standard operating procedures. The knowledge and experience gained from past GIS projects are currently being used to develop new and comprehensive plans for a fully integrated Land Information System (LIS) that will provide managers with automated resource analyses and decision support capabilities.*

## Introduction

America's "frontier", the vastness was glamorized in the movies, neglected because of its emptiness and thought to be worthless. However, less than four decades ago Americans realized that the 440 million acres of the western "frontier" was important — that it contained vast natural resources and wealth, and should be retained in federal ownership.

Thus, in 1946 the Bureau of Land Management (BLM) was created from a merger of the Grazing Service and General Land Office. The BLM is one of the agencies in the United States Department of the Interior. It is responsible for the multiple-use management of the Nation's largest system of federal lands, known as "Public Lands."

Covering nearly one-eighth of the United States, including Alaska, these Public Lands represent all that remains of our Nation's original public domain. Today, more than 270 million acres of BLM-administered Public Lands span the entire American landscape, ranging from tundras to coastlines, alpine meadows to sand dunes, raging rivers to subterranean caves. These lands offer Americans a great diversity and wealth of natural resources.

BLM is mandated by the Congress of the United States to manage the Public Lands under the principles of multiple use and sustained yield. Some of the multiple use programs include: protection and preservation of cultural and historical resources; opening lands for oil and gas, coal, and geothermal development; timber production; fish and wildlife habitat; cadastral surveys; plus a variety of recreation activities, such as development and maintenance of mountain biking and hiking trails, designation of off-highway vehicle and watchable wildlife viewing areas, and promotion of whitewater rafting and scenic driving routes. A vital support function for all programs is lands records data base management.

The majority of Public Lands are located in the western U.S. (plus 30,000 acres managed within 31 states east of the Mississippi River). To bring our offices and employees close to the lands we manage, and in turn better serve our constituents, the BLM administration areas are divided into state offices, district offices, and resource area offices. On-the-ground natural resources program management is developed and monitored by district and resource area offices. It is at these local offices where land management decisions are made, based on sound planning through citizen participation and staff resource specialist recommendations.

The fundamental principles of the BLM lie in the knowledge and mission to wisely manage "land." BLM is responsible to know where the land is — who owns it — what's on it — what's under it — and how to manage it.

Technological advancements that enable managers to make better, more informed decisions about the natural resources within a particular parcel of land, are paramount to successful management. With the increasing complexity of managing resources, the use of technology to facilitate the decision process is essential. Geographic Information Systems (GIS) applications consolidate information previously scattered — from paper maps, files, and memories, and accesses up-to-date spatial data for immediate use.

Continued technological advancements in GIS provide BLM with opportunities to more efficiently research, record, retrieve, update, and interpret pertinent and complex information data bases. GIS is the common thread to facilitate in Public Land administration efforts. The GIS greatly enhances a broad distribution of spatial data to the public and, in turn, facilitates the job of a manager in explaining or defending a decision.

## Land use planning documents

BLM is responsible to determine the best use of the land while accommodating competitive and often conflicting interests, such as wilderness designation versus off-highway vehicle use, or prehistoric cultural sites versus minerals extraction. GIS applications are dynamic. Information must be updated periodically, whenever land status, natural resources, or boundaries change. The BLM's planning process is also dynamic. Comprehensive land use planning is the catalyst of a sound GIS data base.

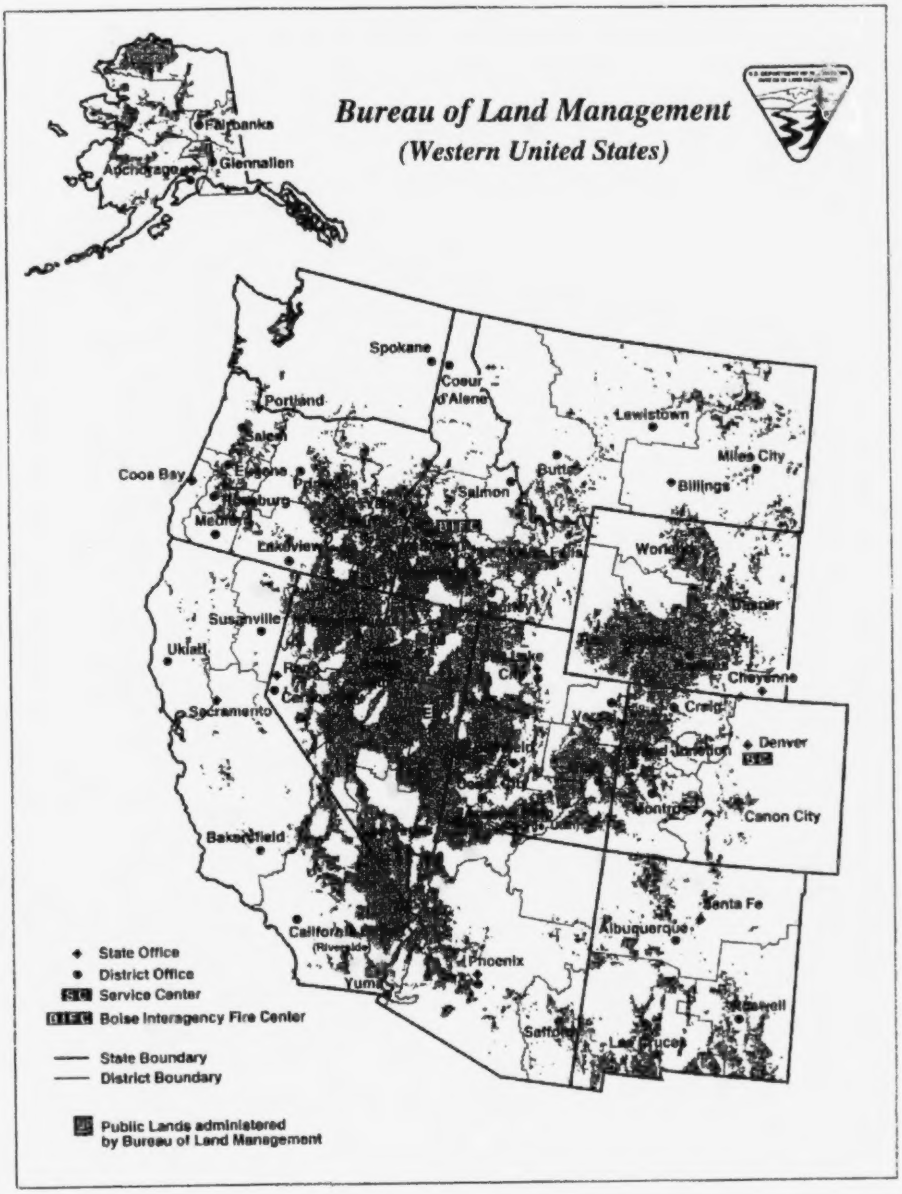
Guided by the National Environmental Policy Act (NEPA), the Federal Land Policy and Management Act of 1976 (FLPMA), and numerous other federal, state, and local laws, BLM balances development with protection of the natural resources to ensure their long-term productivity. In order to do this, BLM develops comprehensive land use plans called Resource Management Plans (RMP's) to set guidelines for multiple-use land management decisions. Land use decisions based on an RMP form the foundation for a data base and reflect all future management actions.

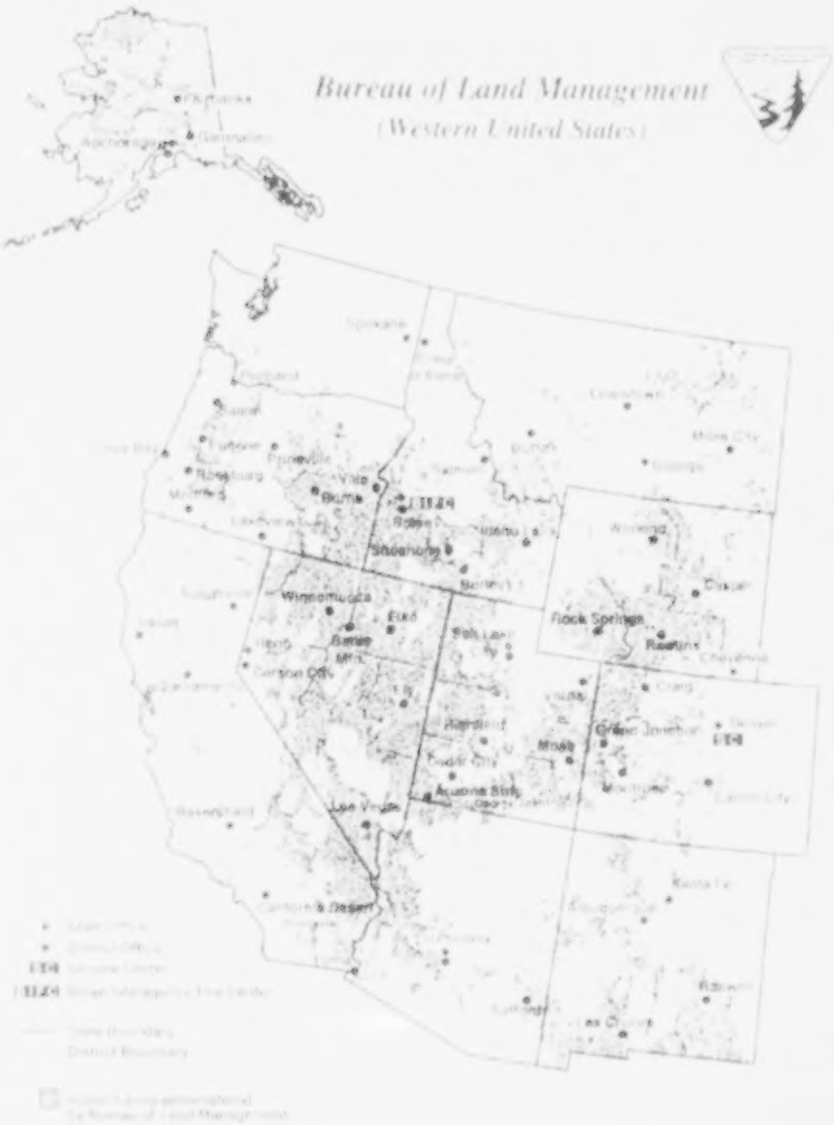
Published in conjunction with each RMP is an Environmental Impact Statement (EIS), which analyzes the effects of implementing each management alternative presented in the RMP. Once all alternatives have been thoroughly analyzed, BLM selects a preferred, multiple use-oriented alternative; one that will best balance uses in an area. These are working documents that guide natural resource managers in complex land use management conflicts. Decisions in the RMP are monitored throughout implementation to ensure that impacts and effects were accurately assessed and goals appropriately met.

GIS-supported RMP's are dynamic — they are constantly updated with new and changing data. Through routine implementation and subsequent plan amendments, there is little if any need to revise planning documents except for very long life cycles or newly emerging issues.

Another important role for GIS is in the development of an activity-specific plan. GIS facilitates the production of site analyses, facility placement, environmental assessments, surface reclamation plans, research projects, predictive models, conflict identification, suitability analyses, weighted composite analyses, acreage calculations, and graphic products among others.

Whether an RMP, EIS, or activity plan, these various planning documents are referred to on a daily basis by field-level managers. They may be prepared for each geographic resource area or individual site, and may vary in size from one acre up to several million acres. Because of topographic, natural, and issue diversity, each plan is unique to the area it addresses. BLM's planning guides are both comprehensive and issue-oriented, and must


Bureau of Land Management
(Western United States)
Fairbanks
Glennallen
Anchorage
Spokane
Coeur d'Alene
Portland
Salem
Eugene
Coos Bay
Roseburg
Medford
Lakeview
Prineville
Salmon
BIFC
Butte
Lewistown
Miles City
Billings
Worland
Casper
Cheyenne
Susanville
Ukiah
Reno
Carson City
Sacramento
Salt Lake City
Craig
Denver
SC
Grand Junction
Montrose
Canon City
Bakersfield
Santa Fe
Albuquerque
California (Riverside)
Phoenix
Yuma
Safford
Roswell
Las Cruces
State Office
District Office
SC Service Center
BIFC Boise Interagency Fire Center
State Boundary
District Boundary
Public Lands administered by Bureau of Land Management


Bureau of Land Management
(Western United States)
Fairbanks
Anchorage
Glennallen
Spokane
Coeur d'Alene
Portland
Salem
Eugene
Prineville
Coos Bay
Roseburg
Medford
Lakeview
Vale
Burns
Lewistown
Miles City
Billings
Butte
Salmon
BIFC
Boise
Idaho Falls
Shoshone
Burley
Worland
Casper
Rock Springs
Rawlins
Cheyenne
Susanville
Ukiah
Winnemucca
Elko
Battle Mtn.
Reno
Carson City
Salt Lake City
Vernal
Craig
Richfield
Moab
Grand Junction
Denver
SC
Cedar City
Arizona Strip
Montrose
Canon City
Las Vegas
Bakersfield
California Desert
Phoenix
Santa Fe
Albuquerque
Safford
Roswell
Las Cruces
State Office
District Office
SC Service Center
BIFC Boise Interagency Fire Center
State Boundary
District Boundary
Public Lands administered by Bureau of Land Management

also include alternative discussions of all affected resources or concerns, which focus on real, on-the-ground problem resolution.

## GIS beginnings

The BLM's first GIS effort began just ten years ago when Colorado embarked on an ambitious planning project that included 603,000 acres of Public Lands in northwestern Colorado. Documentation was needed to be able to quickly analyze alternative development scenarios and to fully address the diverse natural resources with possible social and environmental impacts. The federal government's proposed oil shale leasing program necessitated the development of a comprehensive environmental impact analysis document. This extensive planning project required a thorough analysis of site-specific and cumulative environmental and socio-economic impacts, taking into account the probable development effects of the oil shale resource.

In order to adequately assess most likely impacts, BLM realized that a new and innovative approach was needed to address the complexities of the potential impacts to natural resources associated with proposed land use decisions. The subsequent RMP and companion EIS incorporated acquired information and demonstrated the best multiple-use management direction for oil shale development. The BLM planning process encompassed two years and incorporated GIS in a thorough analysis for a timely completion. The project included digitizing (data entry) and an analysis of complex data using 128 base maps. When final, the RMP contained approximately 900 thematic maps.

Recently, pertinent information about livestock grazing, deer and elk migration patterns, etc., has assisted with management decisions regarding a proposed sodium mine lease, off-site waste disposal, and disposition of nearly 700 oil shale mining claims. Current efforts are now underway to amend and supplement GIS information to evaluate additional Public Lands in northwestern Colorado. Much of the preliminary work is complete, thanks to existing GIS technology.

Early GIS efforts demonstrated that, for efficiency and cost effectiveness and amortization of data collection results in using the data base on a daily basis over the long term. This comprehensive GIS data base is updated whenever necessary, so it does not become outdated. Initial data collected through GIS efforts in the early '80's is continually being updated and used in a variety of planning documents.

## GIS configuration

The GIS used in BLM is the Map Overlay Statistical System (MOSS) family of software developed by the U.S. Fish & Wildlife Service and BLM. The MOSS system includes Automated Digitizing System (ADS) for digitizing, MOSS for vector analyses, and Map Analysis and Processing System (MAPS) for cell (raster) map analysis, and the Cartographic Output System (COS) for finished graphics plotting.

The GIS configuration (both hardware and software) has been standardized throughout the BLM. This configuration will support data standards, and will assist in the implementation of interagency data standards as well. The intent is to systematically improve BLM's current GIS (MOSS family) of software to a specific and limited threshold. Utilizing technological advancements was identified as the highest priority by field employees. The BLM will adopt an incremental approach for making needed improvements. GIS software and hardware will be replaced with a Land Information System (LIS) between 1996 and 1998.

## Current GIS applications

Data requirements to meet planning needs are identified through pre-planning criteria and are limited to those that address key planning issues. If digital information is not available, on-the-ground inventories and/or studies may be conducted and entered into the data base.

To supplement general information in the RMP, specialists enlist the support of GIS "modeling." A map of the geographic area is generated using multiple themes of data (elevation contours, roads, water, land status, mineral leases, etc.). For example, a cultural resources theme could be depicted on a map that also showed the geographic relationship with a number of additional resources also occurring throughout the area. GIS analysis is determined when the overlays are reviewed. Thus, more and more often, GIS is playing an important role in the "modeling" of natural resources information.

Often the analysis and/or combination of data themes provide more valuable information, similar to an inventory but actually an analysis, than the data provided in raw format. For example, if an inventory of elk calving in a particular area was unavailable, it could be modeled and created by combining the soils data with steep slopes, wetland vegetation, recent wildfires, and elevations above 6,500 feet. Acreage summaries and map graphics support the inventories and analyses described in the text. These individual data sets may be combined to derive new maps. Although no physical inventory of the area was conducted by field specialists, the modeling technique using derived data is usually sufficient to project management of the area.

For the past several years BLM's Grand Junction Resource Area Office in west-central Colorado has been assigning resource values to thematic data layers to assist in determining appropriate oil and gas leasing stipulations. Sequential resource conflict analyses assist the manager by quickly focusing on potential conflicts. Resource values can be visually displayed and land use decisions quickly determined. Subsequently, as alternatives are created, they may be displayed as maps maps for immediate use by field personnel.

## GIS in the resource management planning process

Citizen participation and the public scoping process plays an important role in identifying issues and concerns in a specific geographic area. In concert with the physical and biological characteristics of the area, the manager examines resources for their condition, capability, and trends. This analysis is documented in map and narrative formats, thus providing the baseline data base from which alternative management proposals may be developed to clearly document and explain complex management decisions. Current data themes in the data base record the extent of resources, existing conflicts and resource use; provide acreage tables; and produce map graphics that illustrate the analysis.

The next step in the process is to intersect the land ownership theme with each thematic layer. This procedure allows the resource specialists and managers to study the interaction of proposed decisions affecting resources located on Public Lands. The following graphic illustration shows a geographic area in northwest Colorado.

This example of a project area includes roughly 13,000 acres, with several land owners: BLM, Colorado Division of Wildlife, U.S. Forest Service, and private lands. Using the land status as our base map, several natural resource themes are identified. For our example we will use mule deer winter ranges, oil and gas wells, and livestock grazing. Initial scoping meetings are conducted to solicit citizen participation in the identification of issues and concerns. Next, extensive on-the-ground work is conducted by resource specialists. Throughout the entire planning process, thematic data layers are created and updated. The end product depicts a simplified map of the planning area and shows only three resources, but does illustrate how managers must make sometimes difficult decisions about the best management use of a geographic area.

Natural resource planning, as in any other decision making process, requires proposing an array of different alternatives. These alternatives must respond to the issues and concerns as identified in the beginning of the planning process. Each alternative proposed in an RMP represents a complete management plan that can be implemented in its entirety.

If all resources occurring in a planning area are evaluated, it can be quickly determined that some resource uses are compatible with one another, while some are incompatible. Examples might include endangered plants and bald eagles (compatible), cattle grazing and oil and gas development (compatible within limits), and scenic areas and surface coal mining (incompatible).

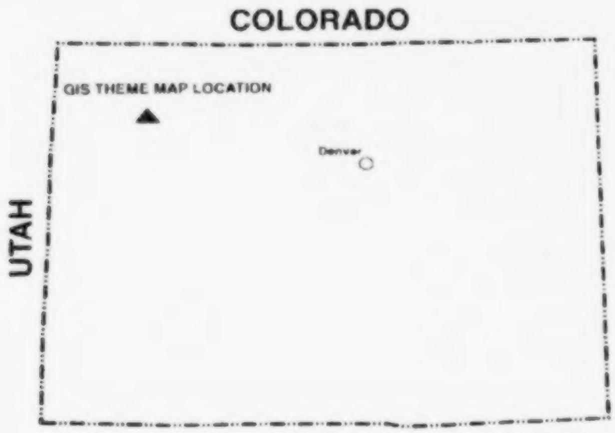
Each resource theme considered in the management scheme may be ranked in some order of preference — that is compared to all other resources to determine the dominant or preferred use or uses within a given geographic area. If two resources existed in the same location, such as coal and waterfowl, the wisest and best use of the land would have to be determined. If the allocation involved compatible resource uses, the decision would be relatively easy. However, in many cases, competing interests are incompatible and must be mitigated.

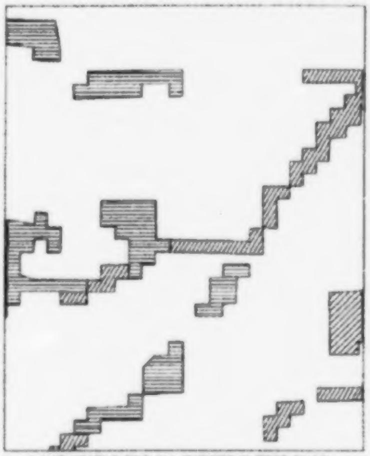
In developing management alternatives, GIS is used to sort areas by compatibility. The theme of each alternative determines the ranking order for its component themes. Components most compatible with the main theme become the highest priority, while those least compatible become the lowest priority in use allocation. Map graphics allow visually portrayed complex data themes. Analysis of the plot relies largely on the ability of resource specialists and other professionals to interpret trends and predict outcomes of actions.

After potential impacts have been identified, a preferred alternative is selected. The alternative is intended to resolve planning issues and promote balanced multiple-use principles with acceptable levels of impact. The preferred alternative can be a combination of various aspects of several alternatives or in some cases might duplicate an original alternative.

GIS cannot and will not replace staff resource specialists professional judgment. Knowledgeable individuals are required to formulate concepts that use GIS applications to interpret the end results of analysis. GIS users are able to analyze large geographic areas with vast amounts of complex resource data in an efficient manner. Analysis, which may be too expensive and time-consuming to perform manually, is often accomplished with GIS capability.

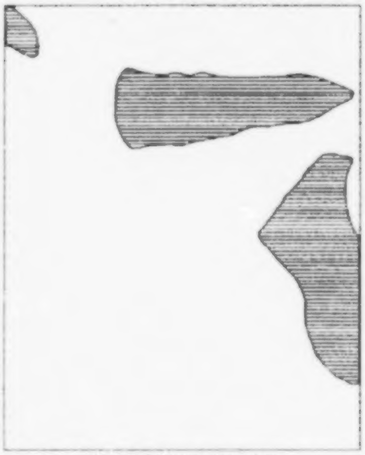
GIS does not and cannot replace the specialist's or manager's ability to quantify impacts, such as the effect clear-cutting timber will have on elk herd numbers, migration, and health. It can help to identify specific conflict areas geographically and quantify the acreage of conflict. For example, biologists know that clear cutting will displace elk herds from part of their traditional feeding areas and reduce desirable cover. They can determine the total habitat loss by intersecting the timber harvest areas with the elk habitat, then create acreage summaries that may tell them that 11 percent of the herd's habitat will be affected. GIS will not identify how last month's record snowfall and cold, or dramatic increases in the numbers of hunters will affect the herd. Biologists will take these additional factors into consideration, calculate the probable effect of habitat reduction, and then determine the cumulative impacts on the herd.


COLORADO
UTAH
GIS THEME MAP LOCATION
Denver



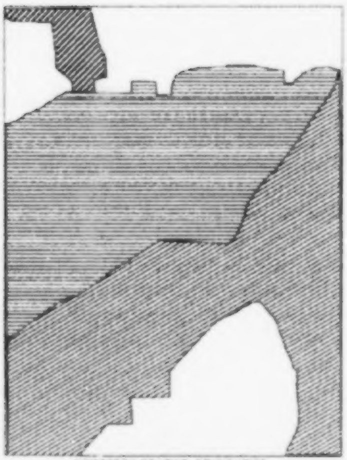
LAND STATUS FOR ALL THEMES IN TEST QUAD

PRIVATE ▤
CDOW ▨
BLM ☐
STATE ▨



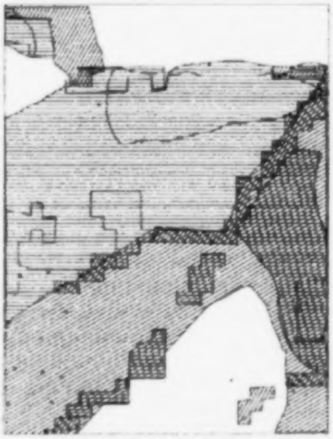
MULE DEER WINTER RANGES

OIL AND GAS WELL SITES



LIVESTOCK GRAZING BOUNDARIES

INDIAN SPRINGS AREA ▨
BOXELDER AREA ▤
SQUARE S AREA ▨



MERGED THEMATIC MAP
(LAND STATUS, LIVESTOCK, GRAZING, MULE DEER AND OIL AND GAS)

In some instances, a management decision may be controversial or challenged by a member of an environmental organization or the oil and gas industry. To defend their position, the manager can graphically show why the decision was made. Ultimately, any necessary changes or additions can be included into the data base thereby keeping all information current and easily retrievable.

GIS serves as a tool to assist in problem resolution and is used to manipulate alternatives, combining various aspects of each. Impacts can be quantified more effectively, producing a higher quality product which can be easily understood by the public, thus reducing the probability of adverse comment or protest.

## Working smarter in wildfire suppression

The following maps show land status with base information such as streams, towns, roads, and transmission lines. For example, in southwestern Colorado, the BLM Fire Management Officer may receive a radio report of a wildfire at a particular geographic point. Using longitude and latitude coordinates of the wildfire, a computer search is generated and plotted. Within minutes fire personnel plot a map that indicates water sources, access roads, potential helicopter hazards with powerlines, and land status. Wildfires do not recognize land boundaries. A lightning strike that ignited an isolated, uninhabited area on BLM lands could easily travel swiftly onto private lands where the possibility of lives and property could be threatened.

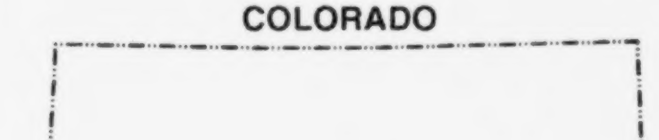
GIS becomes more cost-effective and is amortized when data entered for one project is used in other projects, as indicated in the preceding example. The more that data, equipment, and trained personnel are used, the further the costs, which may initially be high, are spread. This data can be reused for many projects — many of which will be required on short notice without added funding. Some of these include predictive models, conflict identification, acreage calculation, ownership identification, impact analysis, facility placement, and matrix development.

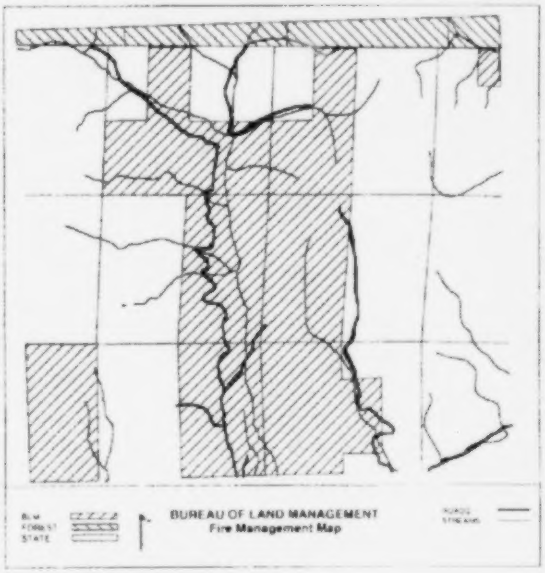
## GIS cartographic advancements

Since most natural resource issues are geographically referenced, they are mappable. The data layers not already on a system, are mapped, then digitized from a preferred U.S. Geological Survey base map at a 1:24 000 scale. Once entered into the MOSS family, the digitized layers are then referred to as resource themes. Plan requirements vary with complexity of natural resource issues but an average of 50 themes is generally needed for most comprehensive RMP's.

GIS automation has greatly facilitated digital cartography as well as updating data already collected. Preparation of custom maps for specific resource themes or for specific activity plans is commonplace throughout all BLM offices. Automated maps that focus on particular uses may be quickly and efficiently produced. GIS-plotted maps are especially useful at citizen participation meetings when a BLM manager needs to illustrate a position or alternative management plan, or enlist public support for a project or decision.

Digital cartography is a costly technological process. The enormous data volume generated by the surveying





sciences creates a demanding work load. The highest costs are typically associated with the data bases. Collecting and maintaining data is an expensive, labor-intensive endeavor. Coupled with the fact that BLM is still in the early stages of developing data standards and consistency, there remain numerous problem areas with graphically correct data output. However, GIS-generated map costs can reduce the manual process costs by as much as 80 percent. Color maps using GIS can be produced in 20 percent of the time required by the manual process.

Although, BLM has partially automated the mapping process with GIS, the agency must still perform some manual tasks for creation of cartographically correct maps. This is especially applicable in the area of final map preparation when four-color printing is required.

## BLM challenges

During the past 10 years, BLM in Colorado has incorporated GIS applications into a variety of natural resource planning documents. In fact, Colorado has been at the forefront in building data bases, and leads the agency in completed RMP's, most of which used GIS.

While success stories continue to increase, there is also a lengthy list of problem areas, stumbling blocks, and resistance to using GIS. Namely, from the very beginning, GIS may have been advertised as a total solution to automating natural resource management applications. However, in some instances GIS efforts have been unrealistic and expectations too high. Learning the MOSS family of software is extremely complex and time-consuming. The lengthy learning process has been interpreted differently by individuals, also knowledge and consistency varies greatly. To become accustomed and comfortable with this technology requires GIS users to "use" the program on a daily basis.

We recognize that no matter how a program is developed, user interpretation will vary. Therefore, probably our greatest drawback to GIS has been user inconsistency. Some individuals in the agency believe that BLM "put the cart before the horse," — that GIS was introduced without the necessary structural changes in work flow and practices that would allow new technology to work efficiently.

Occasionally, due to time constraints, lack of base digital information has delayed or stopped a project or analysis which could have been completed in a GIS environment. BLM is still in the learning curve of data collection. It is not uncommon for personnel to realize the specifics about their area of expertise; however, much of the knowledge resides in their minds and personal notebooks. Subsequently, one of the major goals of automating BLM records to incorporate GIS application is to educate natural resource specialists and managers to collect raw data and ensure that data are appropriately recorded for GIS. The gradual addition of GIS coordinators in all State and District Offices will facilitate and accelerate learning and using the system. We plan to position additional GIS specialists in all BLM offices nationwide.

While GIS coordinators are available to advise and orchestrate data collection and maintenance efforts, it is also each user's responsibility to take "ownership" of acquired data. An automated system is only as good as the data maintained. Unfortunately, data maintenance is tedious and time-consuming. An additional challenge to BLM is to diligently pursue allocating appropriate time and sufficient funding to train personnel. The BLM believes commitment to employee education and training is the most important aspect of adopting the GIS program.

From a manager's viewpoint — many managers do not understand GIS and are therefore reluctant to insist their personnel adopt it. As with any new and innovative program, "old habits die hard." This is perhaps most evident with managers who are not familiar or comfortable with "automation." Rapidly evolving technologies within GIS foster additional apprehension. This is perhaps enough reason to keep our automated efforts simple and user friendly. In a nationwide organization of more than 9,000 people, there will be an enormous variation in GIS interest, knowledge, and acceptance. BLM realizes that GIS applications will provide an improved automated data management system — one that integrates the multitudeof natural resource programs BLM employees deal with on a daily basis. There is no denying the system works — but how do we get there from here? And how can we do it in a timely, cost effective, and efficient manner. The challenge is to prove the system through successful applications. During the last decade BLM in Colorado, as well as early GIS efforts in California, Montana, Wyoming, and New Mexico, provided invaluable evidence of the successes and pitfalls encountered through the GIS learning curve.

As a federal government land managing agency whose budget must be appropriated by the Congress, it is imperative to not only find ways of doing our jobs more efficiently, but to do them at a cost savings too. Ultimately, BLM will be challenged to play a major role in standardizing GIS information for natural resource applications. As mentioned early in this paper, BLM is responsible for the multiple-use management of one-eighth of the land in the United States. Currently, BLM is in the forefront of developing automated data bases. Clearly, GIS meets our required needs and will produce long-term benefits.

BLM has recently initiated Data Standards Teams throughout the field offices. Specifically, these "teams" are organized by natural resource programs, for instance oil and gas, cultural, or forestry. Each team is charged with evaluating the current status of GIS applications, analyzing the information, then recommending agency-wide standards to ensure overall consistency.

Using GIS in the RMP process has created multiple themes of data across the state. There currently are voids in the data base due to collection methods and management priorities. Data themes developed independently by field offices are now being standardized to ensure consistent naming and labeling.

## Summary

What began in the BLM as "humble GIS beginnings" only 10 years ago, have multiplied many times over through the education of our employees, sophistication of GIS applications, simplification of workloads, and increased monetary savings.

As the Federal government's largest multiple-use land-managing agency, it is imperative for BLM to consistently strive for a balanced integrated resource management approach in our decision making. GIS provides us with an enormous and powerful data base. We now must harness this vast abundance of information and methodically transform it into an easily retrievable format.

GIS applications steadily increased throughout the offices in BLM. Many managers quickly realized that GIS was the basis for a dynamic system — a system of unrealized capabilities — and a system that will save time and money.

GIS fosters an interdisciplinary approach and causes people to work together to discuss related resources and conflicts. GIS overlays information and interfaces with natural resources. A wildlife biologist is not alone; dealing only with a specific area of expertise. GIS brings all resource specialists together to discuss current situations and how to evaluate and adjust land use practices for the betterment of not only wildlife, but perhaps recreational uses or enhanced riparian values.

Accurate and reliable land, cultural, and natural resource information is the foundation for successful management decisions. Progressive management strategies are essential to BLM — the Congress of the United States demands it — the American public expects it — and those who are employed by BLM believe in it. GIS is a dynamic multiple-use management tool that continues to reinforce and advance BLM's mission of wisely managing "land."

## References

Dwyer, Michael, Automated Color Separation Mapping, 1986, U.S. Bureau of Land Management, Denver, Colorado.

Zulick, C.A., Application of a Geographic Information System to the Bureau of Land Management's Resource Management Planning Process, 1989, U.S. Bureau of Land Management, Service Center, Denver, Colorado.

# A2 Access to data: grief or relief?

---

**The GRASS CD-ROM global data base and public domain GIS software**

Scott Madry, Rutgers University, New Brunswick, NJ, USA

**Access to data in support of local studies and global investigations**

John E. Estes et al., Department of Geography, University of California, Santa Barbara, CA, USA

**Geographic information system for multidata database management**

Frank Indome, IBM - Bergen Scientific Centre, Bergen, Norway

**Spatial information sharing spells grief or relief ?**

Diane Allen et al., Ministry of Environment, Lands and Parks, Victoria, BC, Canada

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# GRASS CD-ROM global database and public domain GIS software

**Scott Madry**

Cook College Remote Sensing Center
Rutgers University
New Brunswick, NJ 08903-1231
Tel: 908-932-9631 Fax: 908-932-8644

## Abstract

*Global data sets must be made more widely available to the research and teaching communities to support new directions in global scale research. These data must be made available on media that are easily used and must be affordable. The Cook College Remote Sensing Center of Rutgers University and the U.S. Army Corps of Engineers Construction Engineering Research Laboratory (CERL) have signed an agreement that will allow Rutgers to make available on CD-ROM the CERL Global GRASS 1 digital data set. The data set (about 500 Mb) consist of some 50 different raster global data files from various sources; including global elevation, vegetation, green leaf biomass (monthly 1988 AVHRR composites), national boundaries, soils, and marine productivity (Coastal Zone Colour Scanner), among others. Additional CD-ROMs will be developed on an annual basis. This paper describes the global database, the GRASS software package, and their ability to support global research and teaching.*

GIS
92

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Access to data in support of local studies and global investigations

**Joseph H. Bredekamp**

Information Systems Branch
Office of Space Science and Applications
National Aeronautics and Space Administration
600 Independence Avenue
Washington, D.C. 20546
Tel: (202)453-1506 Fax: (202)755-9235

**John E. Estes** (presenting author)

Department of Geography
University of California
Santa Barbara, CA 93106-4060
Tel: 805-893-3649 Fax: 805-893-3703

## Abstract

*Timely, responsive access to data is a crucial element for achieving the science objectives of the U.S. Global Change Research Program and NASA's Mission to Planet Earth. Effective management and utilization of rapidly growing data spatial assets call for new approaches and significant modifications to the infrastructure for accomplishing research. Success will ultimately be measured in terms of the systems responsiveness to science user needs for access to data and the tools required to convert data into new scientific insights. NASA is developing the overall architecture, policies, and implementation guidelines to ensure that data and information systems achieve necessary levels of connectivity and interoperability. A high-level Master Directory will provide open and uniform access to information about space research data. The goal is to provide "one stop shopping" for spatial and other datasets important to a wide variety of investigators. NASA has also initiated an information systems research program to apply advanced computer and information system technology to enhance science user effectiveness. For example, improving methods for: browsing large spatial databases and, visualization of spatial data are being examined. These investigations supported by NASA are intended to improve the overall infrastructure required at scales from local to global.*

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Geographic information system for multidata database management

**Frank Indome**

Department of Information Science,
University of Bergen, HIB, 5008, Bergen, Norway.
IBM-Bergen Environmental Sciences & Solutions Centre,
HIB, Thormølensgate 55, N-5008, Bergen, Norway.

## Abstract

*Database systems (DBS) were a solution to the problem of shared access to heterogeneous files created by multiple autonomous applications in a centralised environment [LITW90]. Globally integrated databases were then used to replace the files for easier data usage. The creation of databases has however led to the collection of different types of data (multidata [1]) for different applications captured using different techniques and at times residing in different formats. And as the world comes closer through computer networking and distributed databases there is the need to integrate some of the data types to support effective decision-making and inferencing that could result from managing multidata. Since traditional database management systems can not handle all data types, this paper looks at multi-data type database systems (MdDBS) and particularly GIS (which is a MdDBS) as a tool for integrating three types of data — facts, statistics and images.*

## Introduction

Current trends indicate that databases in daily use now play a significant role for information analysis and decision-making in both government and private sectors. This can be attributed to the emergence of diverse low priced hardware and software systems, computer networking and a general interest in using computers that have made the process of managing computerised files or databases accessible to a wide variety of users.

The collection and management of data for different applications is common in such areas as business, engineering, medicine, law, education and library science. Traditionally each department or organization collects its information into a database (or file) for a particular application. Different applications however, accept different data types leading to the collection of data from different sources in different formats (types) and at times residing in different applications. The creation of files or databases are thus meant to capture the exact data type needed by the applications used by organisations or departments.

An administrative body like a municipal council, for example, will usually have stored information about individual inhabitants in a relational database, or a list of files and maps showing infrastructure and other related attributes (name of road, length, origin), and hand drawn sketches and aerial images of parts of the municipality. There might also be information showing population census and survey data in statistical formats such as chartsor histograms etc.

In Norway, the medical birth registry collects information on all births and related birth defects and publishes monthly reports which includes statistical tables, histograms and maps indicating graphically the trends in births and birth defects for the municipalities of the country. It stores information in the form of personal information (e.g. mother and child), aggregation of birth defects by e.g. municipality and maps showing changes in rates of some defects by municipality, county defined period of time etc. The information stored is however not integrated in a single system. Flat files are used for some data and a statistical package is used in generating maps for graphical presentations. This example will be referred to as the medical database in this paper.

Typical application data used in the areas mentioned above (e.g. education, offices, etc.) can be classified into five main data types which are represented directly or indirectly in an information system, (in the computing context this means all types of digitally representative data) [DASK91]:

| | |
|---|---|
| **1.Facts:** | Name, address; unique identifier in the types conveniently handled by existing databases. |
| **2.Statistics:** | Data that requires graphical (pie-charts or bar-charts) or tabular (spread sheet) representation ie. the aggregation of the fact data type. |
| **3.Images:** | i.Raster — derived from photo graphs, painting programs, video-dics.<br>ii.Vector — line drawings derived from engineering, computer aided design (CAD) systems, vector graphics.<br>iii.Moving — a sequence of the above with extra clues to 3-dimensional relationships, history, time-series, narrative. e.g. animation |
| **4.Text:** | As in a book or magazine article |
| **5.Sound:** | E.g. the human voice — narrations, comments, lectures, music and sound from nature. |

Collections of different data types are termed multidata and information systems that integrate *at least two* different data types are termed **multidata database systems (MdDBS).**

Such an integration, for this paper, entails bringing together the different data types within a single coherent system such that users can have a unified view of their data.

Numerous advantages accrue if multidata collected by organizations can be effectively integrated within a single information system [SHEP91]:

- reduced costs associated with data acquisition through the elimination of duplications in collected data and conversion activities (ie. from one data format to the other)
- more operations can be performed on an integrated data set than on disparate sets of data
- sharing of information by organisations leading to more cooperation
- better presentation of ideas and products to clients and the public
- effective decision-making
- better inferencing from collected data
- perceived seamless information environment by users.
- quicker retrieval of multidata

Traditional database management systems (DBMS) have been designed to handle facts only. The integration within these systems involves storing the fact data in single databases with no redundancy. They however do not create the environment for integrating multidata as they lack facilities for handling other data types. Their relational operations, for example, are not directly applicable to some data types.

Aggregation functions, for example, cannot be applied directly in cases of missing sample values. Some types of data may be collected or measured using various scales, measuring distances or coordinates which complicate the integration of the data.

Traditional DBMS in general do not have graphical capabilities. Histograms and pie-charts and three-dimensional visualization, as examples, cannot be created directly. They are also not particularly suitable for spatial (i.e. image data) applications because they do not efficiently support the type of operations required for graphical applications.

There is a great deal of research taking place with diverse ideas about the integration of multidata - some developments are in the prototyping and others might take some time before being marketed commercially. The problem for this paper then is, how can organizations integrate their multidata at this present moment? It is the belief of this paper that some existing systems can be used for such an integration. One such system is the **geographic information system (GIS)**. GIS are DBMSs known to manage at least two types of data: fact and image data (vectors and rasters), and can thus be classified as MdDBS.

In the sections that follow a reference model for an ideal MdDBS is presented, the requirements for management of multidata are given and how a GIS could be used in integrating multidata is also presented. An assumption made is that the data types could have been collected for other applications and not specifically for a GIS.

## Multidata DBS (MdDBS)[2]

A database system (DBS) has been defined to be the set of data and some software for managing the data called a DBMS [SHET90]. The data referred to here represents the fact datatype. For the purposes of this paper this definition is extended to a set of multidata and a DBMS. This extended definition then refers to a multi-data (type) DBS (MdDBS).

The focus of database research and development thas been more on databases that store fact data (structured data) with the conceptual representation of the data provided through the DBMS. Logical concepts like objects, their properties and their inter-relationships are used in the representation and users need not know details about how the data is stored physically.

MdDBS use similar concepts as above. They mainly differ from conventional DBMSs by their ability to handle more than one data type. The fundamental basis for the design of MdDBS is also based on the techniques used by traditional DBMSs. A MdDBS thus maintains knowledge about the underlying storage structure of different data types, and the ability to manipulate the actual database or files involved. A good database design therefore, needs to be supported for efficient and effective management of all the data types within the MdDBS. For example it should be possible to process, in an interactive manner, fact data into aggregates (i.e. statistical data) and present results in different formats (i.e. pie-charts, bar-charts or maps).

The reference model presented below tries to implement an efficient and effective integration of multidata. The model is an extension of that for a centralised DBS [SHET90] and combines data/schema integration with query processing through the following basic components:

- **data:** Facts, statistics, images i.e. vectors and rasters
- **queries:** Requests (or commands) for specific actions that are either entered by a user or generated by a processor
- **processors:** Programs or modules that manipulate commands and data
- **schemas:** Descriptions of data managed by the DBMS and consisting of objects and their inter-relationships.

There are three main processors in this model - accessing processor, transforming processor and the interpretation processor.

- **Interpretation processor:** This processor reconstructs commands and associated data that can be sent to another processor. It also serves to partition and /or replicate an operation from one processor into operations accepted by multiple processors. It decomposes queries and checks for syntactic/semantic constraints.
- **Transforming processor:** Translates commands that are expressed using the conceptual schema objects into commands using the integrated schema objects. This provides a type of data independence by which the data structures and commands used by one processor are hidden from the other processors e.g. transforming data from one format (image data) to another (fact or statistical data).
- **Accessing processor:** Executes commands against the database. This access can be based on different methods eg. access to data based on value (i.e. name, building, address, county) or spatial location (by coordinates, latitude and longitude). It may accept commands from several processors and reshedule the processing of the commands. The functions of the accessing processor include:
    - The management of the file system and access to the data dictionary
    - Generating data required by a user through queries or to the processor generating the query or command

There are three schema levels which portray the levels of abstraction — external, conceptual and internal schemas.

- **External schema:** Concerned with the way the different data types are viewed by users. This forms the subset of the database that is accessible by a group of users
- **Conceptual schema:** Describing the logical data structures (i.e. the schema consist of objects that provide a conceptual or logical-level description of the particular data type) and the relationships among the structures
- **Internal schema:** Describing the physical characteristics of the logical data structures in the conceptual schema e.g. information about the placement of records on physical storage devices.

The reference model has a design that aims at the efficient integration and management of facts, statistics and images. The different external schemas are a way of having many users access the database and at the same time limit access to sensitive information. Some users, for example, may be able to access the whole database whilst others might have access to only statistical data (i.e. the aggregation of the fact data).

Even though users need shared access to the different data types their needs might be different. One user may need to access only images whilst another may want to manipulate facts into statistics then produce a graphical representation of the results through maps, charts or histograms. Filters are used by the interpretation processor for these multiple user views of the database.

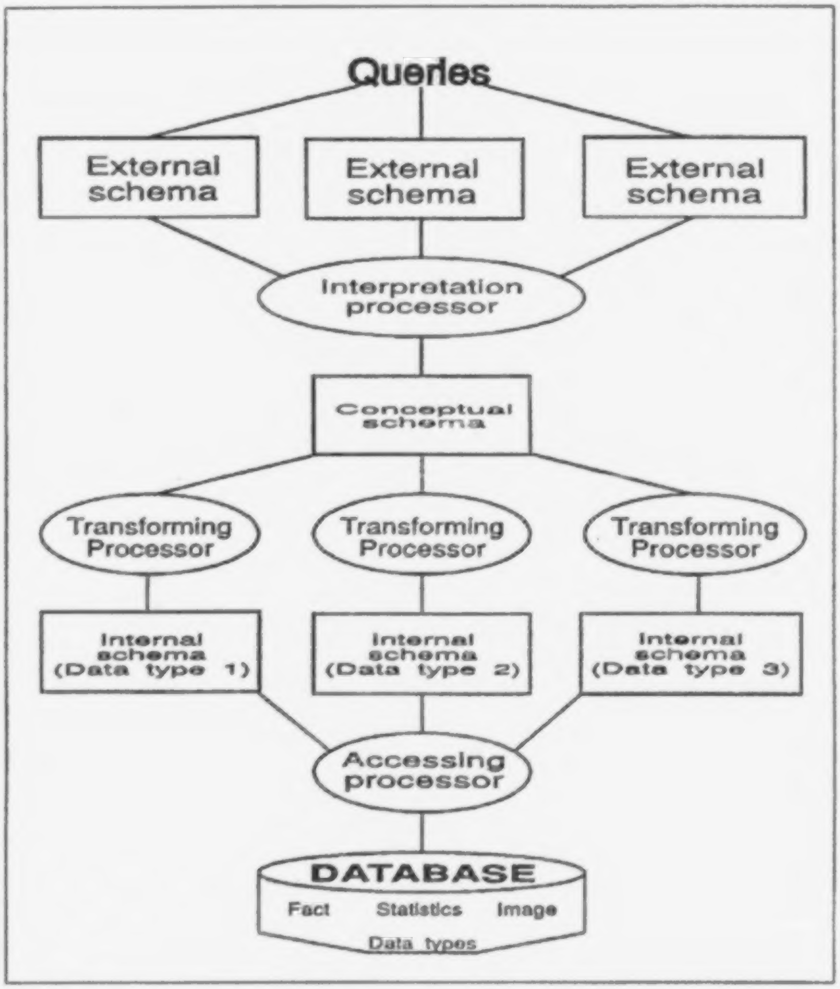

Queries
External schema
External schema
External schema
Interpretation processor
Conceptual schema
Transforming Processor
Transforming Processor
Transforming Processor
Internal schema (Data type 1)
Internal schema (Data type 2)
Internal schema (Data type 3)
Accessing processor
DATABASE
Fact
Statistics
Image
Data types

**Figure 1:** *Reference model of MdDBS*

The multiple internal schemas are quite unique for this model. The main reason behind these is to allow the system take advantage of the best possible data model for each of the data types. Thus a complete MdDBS could comprise of two data models — a relational model for the facts and statistical data types and a network model for the images (i.e. a raster representation).

Relational database models are well developed for fact data but are not that efficient when used to represent image data types e.g. vectors or rasters. In the case of vectors, large amounts of data ought to be processed before features and maps can be drawn resulting in poor response time with querying. Data models can be developed e.g. through a hybrid of some existing models for better representation of e.g. image data. The multiple internal schemas, when they are all relational, however make this model resemble a centralized DBMS.

The representation of the different data types are then transformed to an integrated conceptual schema by the transforming processor. The accessing processor serves as an interface with the operating system for faster storage and retrieval of the data requested. This helps reduce the number of physical accesses necessary to retrieve data for a particular process e.g. drawings or conversions from one data format to the other.

## Requirements of MdDBS

The overall purpose of MdDBS, like other information systems, is to maintain information and make that information available on demand. In archiving this aim the system is organised by functions that can be classified as follows:

a. input — how the system takes in the facts, statistics and images data types
b. storage — how it stores these data types
c. retrieval — through querying and browsing
d. manipulation and analysis — generation of new types, updates, expansions
e. presentation — output of results

### Input

A MdDBS should be able to integrate both new and old (i.e. already collected) data. There should be facilities that enable users transfer data into the system. This should include e.g. input through keyboards and transfer of files in particular formats for the different data types; scanners and digitizers for images.

### Storage

A lot of research has gone into the representation of fact datatype and at the moment there are many databases that manage this datatype, information gathered can be stored as relational tables. The MdDBS should thus be able to handle the data structures and support a query language able to search for attributes based on the primary key system.

A statistical database can be seen as a set of tables and a set of derivation functions; each function maps a set of tables into a new one [BARC88]. This statistical data type consist of statistical or aggregate data used for statistical analysis e.g. studies with socio-economic trends, energy production and consumption, health and medical statistics, population and income counts.

Due to the nature of this datatype some research efforts call for a new data structure different from that of the fact datatype. For example, most of the data models for conventional database support merely "simple data structures" generally represented by "relations" whereas statistical data types need to be supported through "complex data structures" generally expressed in the form of tables (time-series, data matrices, contingency tables) [RAFA88].

The system should also be able to support storage of the different spatial (image) data models. These models store the image data in digital form i.e. vector and tessellation models. The vector type uses topological elements (points, lines and polygons) and the set of interconnected coordinates of one or a combination of the elements that are linked to the facts. The tessellation model comprises of any infinitely repeated pattern of a regular polygon and is referred to as a raster or grid.

### Retrieval

Retrievals in the form of queries should be possible for each data set and a combination of the data sets through Boolean logic e.g. join, semi-join etc. A user should be able to browse through a list of options and then retrieve information based on some of the data sets. An example from the medical database is finding personal information of all babies born with a particular birth defect.

With reference to the model of section 2, retrievals should be based on the type of external schema available to the user. Through filters, different users would have different information that can be retrieved from the system. There should also be retrievals by data type or data structure which are different from retrieval by content, as above. An example will be finding all information stored as images.

### Manipulation and analysis

The following operations and functions should be supported for the data types in the system:

- set operations: unions, intersections, difference e.g. tables, maps, overlays
- logical operations: AND, OR, XOR, NOT
- search operations: finding attributes, images in the database
- similarity operations: translations, rotations, scanning
- graphical analysis: polygon overlay, route allocation, intervisibility
- statistical analysis: histograms, frequency analysis, measures of dispersion
- measurements: lengths, areas, volume
- time series analysis: temporal data analysis
- modelling capabilities
- structure conversion: between different models and input formats
- geometric conversions: scale changes, map projections
- generalizations and classifications
- enhancements e.g. for images
- abstractions e.g calculation of image centriods
- automatic updates: update of the different data sets
- interpolation and extrapolation functions: for missing data

### Presentation

Output from the system in the form of presentations should include the following:

- graphical displays of tables, maps, charts, histograms etc.
- hard copy formats of the above
- tabular displays
- reports of facts and/or statistics generated in an interactive mode

## Integration with GIS[3]

The techniques (e.g. storage, retrieval) that are used in the traditional database systems form the fundamental basis for the design of GIS. GIS is a database system that supports the storage, analysis and manipulation of geographic data and the display of data in graphical form (maps, charts, histograms). It is an information system with the ability to relate the position and relationships of objects on the earth surface and describes the collection of geographical objects over some n-dimensional map with stored information describing entities that have a physical location and extent in some spatial region of interest.

Geographic objects can generally be classified as belonging to a particular entity class such as a town, road, lake etc. These objects are described by their associated alphanumeric attributes which are termed aspatial (e.g. age,name etc.) and their spatial attributes (e.g. location in terms of projection coordinates, latitude and longitudes). Queries within a GIS involve identification of the entities belonging to the objects based on their aspatial and spatial attributes and the spatial relationships between the entities.

The idea behind using a GIS to integrate different types of data especially facts, statistics and images not collected specifically for a GIS is based on the data types that the latter manages which can be grouped into the following categories [BENG90]:

**Aspatial data:** This consist of information about an object (e.g. a house, a well etc.) i.e. the alphanumeric attribute-based data normally stored in traditional database systems. e.g. name, identification number, address, population. Such data is stored as tables or relations. Most GIS use traditional RDBMS to manage this data type

**Spatial data:** Made up of information about the position and location of the object. This is in the form of coordinates and other spatial properties describing the spatial boundary of objects whose corresponding descriptions are stored as aspatial relations. A relational database system is usually used for the storage of such data.

**Graphical data:** Graphical data are for displaying results and other outputs from the system and can be in the form of maps, n-dimensional graphs and charts. Normally the spatial data is defined within the graphical (map) space. The graphical data is thus closely related to the spatial attributes of the spatial data.

A close look at these data types shows some similarities with the data types defined in section 1 (i.e. facts, statistics, images, etc.). The aspatial data is actually the fact data type. This forms the structured data stored in traditional DBMSs. When the aspatial data is summarised by aggregation functions they become statistical data and by using a graphical output facility the latter can be presented in forms such as charts, histograms etc. The spatial data is partly of facts and partly of the image data type. The locational attributes are image data and their corresponding descriptions are stored as facts. Finally the graphical data for presentations are of the image data type.

The integration of data using a GIS is seen as the synthesis of geographical information in a computer system which depends on its effectiveness on information linkage (i.e. of spatial and other data types) within a coherent data model. This involves bringing together diverse information from a variety of sources, requiring effective matching of supposedly similar entities in these sources and demands information consistency across the data model [SHEP91].

The integrated data sets are represented as maps or layers and a GIS maintains a hierarchic object view of the data. The layers can be manipulated by functions such as addition, subtraction and searching for patterns. There are query facilities through which the hierarchic data sets can be addressed. Diverse questions can be asked about the data in a GIS. [MAGU91] lists six basic questions based on [RHIN90] which are given below with examples from the medical database.

1. Location: *What is at ...?* For finding what can be found at a location. This query can be used to retrieve information based on particular datatypes or a combination of them. e.g. How many births were recorded at municipality X last year? List name and address of mothers who delivered at county Y in Spring of last year.
2. Condition: *Where is it ...?* For finding characteristics at a location. e.g. Where is the rate of birth defect Y greater than 200 per 10,000?
3. Trend: *What has changed ...?* For checking changes that occur or might have occurred for a location over some period. e.g. What was the change in rate of birth defect Y for 1990 for county B?
4. Pattern: *What is the pattern ...?* For describing and comparing the distribution of phenomena. e.g. Is there a correlation of an increase in birth rate with decreasing iron concentration in drinking water?

5. Routing: *Which is the best way ...?* For finding the best, fastest, quickest etc. route to a place. e.g. Which is the shortest path to a hospital?
6. Modelling: *What if ...?* For evaluating models based on defined criteria. e.g. How will the rate of birth defect X change if zinc concentration in drinking water increases by 15%?

Using a GIS to integrate multidata the main requirement will then be geographic data i.e. data showing a spatial region of interest, which will be used to link or connect the data types being integrated to a specific location on the earth. This will not be difficult since information or data collected can directly or indirectly be related to a specific place.

As an example an existing university database can be spatially referenced by the location of the university in a town, city, municipality or even in a particular country. The locations of the municipalities or counties in the medical database can form the geographic information for an integration using a GIS.

Commercial information systems integrate fact data by storing them in a database with no data redundancy. Mapping systems also integrate spatial data within single graphical databases i.e. in either vector or raster format. Answers to questions or queries to these systems are based on the relationships between items of consistently recorded data, with the range of questions being of a spatial or fact nature.

Within GIS, however, the linkages between aspatial and spatial data describe real world features and the operations performed, in addition to those above, are across multiple sets of data (e.g. between aspatial, spatial and graphics) producing a diverse set of questions that may be asked and a broader range of problems can be solved than in systems that handle only one data type[SHEP91] .e.g:

- different items of facts data can be associated with one another through a shared location
- users can interrogate geographical features displayed in a computer map and retrieve associated information for display (i.e. in tables or charts) or for an analysis
- maps can be constructed by querying or analysing attribute information in a database
- new sets of information can be generated by performing spatial or aspatial operations e.g. overlays of polygons or the aggregation of facts into statistics.

The linkages which are the key to the integration of data types within a GIS differ in technology from one GIS to the other. The basic technique, however, involves establishing pointers between each spatial feature in the database and the associated attribute (aspatial or fact ) information.

The task for integrating already collected data will then involve the creation of those linkages between the geographical information which will be stored in the GIS and the files or databases storing the other data types. Sometimes geographic data in form of vector coordinates might exist but not specifically for a GIS. The transformation of such data for a GIS sometimes involves defining geo-reference points.

## Conclusion

Integrating multidata stored in files or databases has been an important area since the inception of information management due to the different applications used for collecting and storing data. The past few years has seen considerable progress made in the use of GIS to integrate different types of data.

A GIS does not fully meet the reference model of section but it does manage facts, statistics and images (rasters and vectors). [HEAL91] describes two different approaches to GIS design meant for this integration. The first termed the hybrid model uses linkages to connect attribute tables in a database and a set of direct access operating system files whilst in the second model the GIS serves as the query processor sitting on top of the database itself. They all aim at the best method to interlink diverse types of data to form an integrated information system.

For now, organizations wanting to integrate different datatypes can rely on GIS as the latter has the capabilities of a graphical interface and support of a database management system which does integrate diverse data types.

## References

[BARC88] Barcaroli G., Di Battista G., Fortunato E. & Leoporelli C.Design of statistical information media: Time performances and storage constraints. In: IVSSDBM. Statistical and Scientific database management M. Rafanelli, J.C. Klensin & P. Svensson (eds). Lecture Notes in Computer Science. Vol 339. Springer-Verlag, Berlin pp.1-3

[BENG90] Beng Chin Ooi, 1990 Efficient query processing in geographic information systems. In: Lecture Notes in Computer Science Vol. 471, Springer-Verlag, Berlin ,1990.

[BURR86] Burrough P.A., 1986 Principles of Geographic Information Systems for Land Resources Assessment. Clarendon Press, Oxford.

[DANG90] Dangermond J., 1989 The Organizational impact of GIS technology. ARC News Summer:25-6

[DASK91] Daskalakis A. & O'Docherty M.H., 1991 Multimedia Information Systems in the Management and Semantic Retrieval of all Electronic data Types. In: The Computer Journal Vol.34, No.3, 1991.

[FRAN88] Frank A. U., 1988 Requirements for a database management system for a GIS. In: Photogrammetric Engineering and Remote Sensing 54 (11): 1557-64

[HEAL91] Healey R. G., 1991 Database Management Systems. In: Maguire D.J., Goodchild M.F. Rhind D.W. (eds) Geographical Information Systems: principles and applications. Longman, London pp.260-263.

[LITW90] Litwin W., Mark L. & Roussopolos N., 1990 Interoperability of Multiple Autonomous Databases. In: ACM Computing Surveys Vol 22, No.3 Sept.1990

MAGU91] Maquire D. J., 1991 An Overview and Definition of GIS. In: Maguire D.J., Goodchild M.F., Rhind D.W. (eds) Geographic Information Systems : principles and applications. Longman, London pp. 9-11

[RAFA88] Rafanelli Maurizio, 1988 Research topics in Statistical and Scientific database management: IVSSDBM. Statistical and Scientific database management M. Rafanelli, J.C. Klensin & P. Svensson (eds). Lecture Notes in Computer Science. Vol 339. Springer-Verlag, Berlin pp.1-3 tion Yearbook 1989. Taylor & Francis and Miles Arnold, London, pp.85-91.

[SHEP91] Shepherd I.D.H., 1991 Information Integration and GIS. In: Maguire D.J., Goodchild M.F., Rhind D.W. (eds) Geographic Information Systems: principles and applications. Longman, London

[STER85] Stern R.A & Stern N., 1985 An Introduction to Computers and Processing. 2nd. Edition. John Wiley and Sons: New York pp.618-625

SHET90] Sheth, A.S. & Larson, J.A., 1990 Federated Database Systems for managing Distributed Heterogeneous and Autonomous Databases. In: ACM Computing Surveys Vol.22 No.3 Sept 1990

[1] *Multidata and multi-datatype refer to different types of data*

[2] *An information system is designed so that it can be accessed by both management and operations personnel for producing required results and for assisting in the decision making process '[STER85]*

[3] *A GIS brings together information, it unifies and integrates that information. It makes available information to which no one had access before, and places old information in a new context. It often brings together information which either was not or could not be brought together previously' [DANG89]*

# Spatial information sharing spells grief or relief?

**Dianne E. Allen P.Eng. (presenting author)**
**Hally D. Hofmeyr P.Ag.**

GLIDE Unit
B.C. Ministry of Lands and Parks
4th Floor - 1802 Douglas Street
Victoria, B.C.
V8V 1X4
Tel: 604-387-8451
Fax: 604-356-7831

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

## Abstract

*We are entering a new era of land-related information use. It has become evident that the sharing of our data resources is essential for sound decision making based on the most complete and most timely information available. Whether for integrated resource management or municipal management this means that information from various sources and disciplines be shared and combined.*

*There is anticipation of easy access to data, more complete analyses, lowered costs and reduced duplication of effort. But there is more to data sharing than simply data access. And spatial data is more difficult to share than data that is not spatially referenced. This paper will explore impacts on the users and providers of data. The way we manage data will change and with the increased importance of data administration so will our skill requirements. Even organizational structures and the way we carry out our business may be affected.*

## Introduction

We are entering a new era of land-related information use. Governments and the private sector are now viewing data as a corporate resource to be shared within the organization or between affiliated organizations. There is growing recognition that the cost of data collection for a new initiative greatly exceeds the costs of hardware and software acquisition and that data very often has a value outside of its original application. It has also become evident that sharing of our data resources is essential for sound decision making based on the most complete and most timely information available. Spatial data, as used in Geographic Information Systems (GIS), can be merged from different sources to give a multi-disciplinary view of a location that is so important to integrated resource management and municipal planning.

In British Columbia the Land Information Infrastructure (LII) (Hofmeyr et al., 1991) is being built to facilitate the exchange and sharing of land-related data between provincial government ministries. An operational prototype will be delivered in the spring of 1993. There is anticipation of easy access to integratable data, more complete analyses, lower costs and reduced duplication of effort. But there is more to data sharing than simply data access. And spatial data is more difficult to share than data that is not spatially referenced. Understandably some participants in the LII are proceeding cautiously.

This paper will give an overview of expected impacts of data sharing on the users and providers of data. An understanding of these impacts and the problems they may create is essential for participants to take appropriate action at stages in the implementation of a data sharing mechanism such as the LII. Upon such actions the success of data sharing initiatives depends.

## Technical requirements for data sharing

The technical problems needing resolution before data sharing can become an accepted business practice are not the focus of this paper. They will be reviewed however for completeness. Noting how they are addressed in the design of the LII will provide a context for the discussion on impacts of data sharing.

The B.C. Government's land-related data sources are distributed widely among many ministries and in physically disparate locations across the province. To accommodate this the architecture of the target LII is a heterogeneous multi-database management system (MacDonald Dettwiler, 1991). The actual transmission of data between sources and users is thus an issue. Land-related data sets are typically large volume files. Speed of transmission offered by communications networks is a limiting factor in the practicality of on-line access to distributed data sources. Plans are to use the existing Municipal Area Network in Victoria and expand the on-line access to the rest of the province as government telecommunications capabilities improve.

Data sharing would be frustrating if users had to learn the idiosyncratic access methods of each database. A common access method is preferred and is to be provided in the LII by the User Access Terminal specifications. Through the User Access Terminal users will be able to browse, order and receive data.

Facilitating the exchange of data will be a Data Access and Management Facility consisting of a number of task servers. Even with data access improvements the level of data sharing and exchange will not increase unless users have an easy way of finding out what data is available. Linked to the Data Access and Management Facility will be the LII Data Directory, providing an organized list of land-related data sources. A graphical index to B.C. is important for locating the appropriate spatial data. A common understanding of the data will be captured in a Common Data Model which will provide a semantically consistent view of all the data accessible via the LII. Closely linked to this will be a Data Dictionary which defines the data and assists users in accessing it.

The most effective way to manage land-related information is still an issue. Many of the B.C. Government's data sets are being migrated to new database management systems to improve access and analysis functionality. Managing attribute data plus the associated spatial data continues to challenge custodians. The architecture of the LII will allow the database management systems to evolve as is appropriate for the particular data sets.

The variety of data formats in use on GIS platforms has been a significant impediment to data sharing. It would be impossible to demand that all databases across the B.C. Government use the same data format. However a common exchange format would simplify the translation requirements between ministry systems. Data will need to be reformatted for transmission to users. The exchange format must be carefully chosen so as to avoid the loss of data or data relationships upon translation. The data exchange standard for the LII will be compatible with the National Standard for Canada (B.C. Government Land Information Data Exchange Unit, 1991).

Likewise a common georeferencing system is required if spatial data from various sources is to be combined. In B.C. the NAD83 datum forms the standard positional framework (B.C. Government Land Information Data Exchange Unit, 1991).

A multi-agency management structure is required to develop, recommend and institute standards and policies for the operation of the exchange mechanism and for sharing of data. In the B.C. Government the Land Information Strategic Committee, with representation from key land information providers and users, supported by the Government Land Information Data Exchange Unit, plays this role.

## Considerations for users of external data sources

The promise of easy access to many data sources, without the hassles of reformatting and coordinate transformation sounds utopian to many decision makers. However *caveat emptor* applies here. Knowing that data from another agency was captured for a particular purpose users need to consider if it is appropriate for their own application. Accuracy, reliability, currency and volatility need to be assessed as well as the data structure. Data coverage, cost and copyright will also determine its suitability. So the users cannot accept data from external sources without some investigations.

Since digital data can be so easily rescaled for display and overlay, some users will be tempted to ignore the intended scale of data. However the accuracy of the data, both positional and taxonomic, will be influenced by the intended scale of display of the data and the resulting generalizations that were made. It will also be influenced by the scale or precision of data capture. The processing history may also give some clues as to the data accuracy.

The reliability of the data can be gauged by considering how the data was captured and by whom.

Its currency will depend on its date of capture and the changes on the ground that may have occurred since then. Will the data be too outdated for the user's purposes? The frequency of updates to the source database will need to be considered, and whether a "snapshot" of the database is all that is needed or if updates need to be incorporated on a regular basis.

The data structure, i.e., whether it is topologically structured or simply spaghetti line work, will determine whether the data can be used in different types of spatial analysis.

Some data sets do not have complete coverage of the province or it is in varying scales or of drastically varying ages.

The cost of the data is a most important consideration as are any copyright restrictions that may limit the use of the data and the marketing of value added products.

Even with an exchange mechanism such as the LII in place, there will continue to be restrictions placed on access to some government databases due to security, particularly privacy, concerns. Data custodians will still make decisions as to who may access their data. So, potential users may need to negotiate special arrangements for data sharing. Also, if there are no data sources that meet users' requirements they may be able to enter into cost-share agreements with an agency involved in data capture, to arrange for the needed coverage or accuracy.

It is evident from the above that a thorough investigation of potential data sources needs to be done. Confirming data sources is best done early in the planning process of any venture since the cost of data capture or acquisition is potentially a major budget item and since special arrangements with data providers may need to be done well in advance.

## Considerations for data providers

Generally, the decision to share data with outside users is made by the data custodians themselves. There may be instances when the data is deemed at the overlying corporate level to be of value to the entire organization and pressure is put on the data custodians to release their data. Whichever the case the result will be changes in the day to day operations and the program of the data custodian, now a data provider, and perhaps changes in the organizational structure.

Data administration and database administration functions within organizations will become more critical and have a higher profile as agencies endeavour to meet corporate-wide standards for data management, such as data modelling and provision of metadata. The necessary expertise may need to be recruited or the present staff trained in data management. The trend, in both the private and public sector, towards viewing data administration as a management function rather than a technical function will likely continue (Durell, W., 1985). As data exchange becomes a frequent interaction between agencies, organizational excellence may even be judged on the basis of the strength of data management functions.

As more potential users become aware of data sources through the central Data Directory and as access and data conversion is simplified, data custodians can expect a greater demand for access to their data. And as awareness increases, there will likely be a demand for equitable access facilities regionally across the jurisdiction, and pressure for access by the public. Once established these will continue to increase data usage.

Security of the data is an important issue when on-line access is being considered. This includes protection of privacy and the protection of data integrity. Data custodians will have to decide what data cannot be released to which users because it contains personal or other information as defined in provincial legislation (Leia, G.J., 1989). How one can control the distribution of data once it has left the custodians control is still a concern (Kozub, N.E., 1991). Protection from unauthorized editing of databases should be provided by the access mechanism.

On-line access to data will minimize program disruption in the providing agency as staff will not be required to cut and ship tapes, etc. In the case of the LII, access and accounting will be handled outside of the providing agency. However, the priority given to LII transactions over internal transactions will need to be determined.

A drawback to on-line access is the lack of negotiation, discussion and education that would likely have taken place when data access had to be arranged on a case by case basis. Data custodians may feel that their authority and control over their data has been usurped. And they may worry about the misuse of their data. By including quality information along with the data, custodians will enable users to make informed decisions regarding the suitability of data for their use. Participants in the LII will be submitting metadata to the Data Directory, such as source, measures of accuracy and reliability, date of collection and processing history. But there is no guarantee that users will consider that information or even understand its implications.

This raises the question of what is owed the users of the data for which one is responsible. Disclaimers as to responsibility for damages arising out of use of the data may not hold up in a court of law. Providers owe users a duty of care. Legal liabilities will need to be defined for all roles: the data provider, the data transmitter, and the data user.

In the event of a liability dispute will the provider be able to prove that data was altered by a user? This would require regenerating a snapshot of the database at any point in time, necessitating careful planning and control of data backup and archive (Guthrie, J. 1991).

Already mentioned is the need for a common data model, to ensure a common understanding of the data being shared. But whose understanding will be reflected here? Likely it will be the data custodians', but will this be meaningful to potential users? Will data providers be obliged to educate users. Paper maps often contain or are accompanied by documentation on map use. Many people acquiring maps at a government or private dealership or at a map library do not know what they really require and are assisted in their search by an expert in map use. The task of educating digital map users is more complex, given that users will be able to "serve themselves" using on-line access and given the flexibility of the data and resulting ease to which it can be used, and misused, in all sorts of automated analyses.

Will users be registered and sent notices of data updates? Conventional map suppliers do not automatically send updated maps to their customers, but they do often send annual catalogues which list updates and new products. Perhaps these same procedure can be followed for digital data.

Corporate wide standards may oblige agencies to collect and manage data to standards which are more stringent than those required by their own programs. The added costs may push them to seek cost sharing arrangements and coordinate with other agencies.

The new interdependence between agencies, or departments, will mean the need for more communication and teamwork (Edmondson, P.H., 1991). Separate disciplines will need to interact and gain a common understanding of each others needs.

## Mitigating the grief; reaping the relief

Agencies may want to delay their participation in data sharing initiatives like the LII and instead wait to see what costs are incurred by participants and whether the advantages outweigh those costs. However participation at the outset will ensure that an agency can begin planning to take best advantage of the exchange mechanism and can influence the endeavour to best meet their own needs.

An agency may decide to wait for the data sharing mechanism to be running smoothly before participating, only to find at that time that they do not have in place the internal procedures or expertise to make use of it. As well as there being a learning curve to consider when implementing new procedures, there is a familiarization curve involved in instituting a new way of doing business. Data sharing can impact program planning, budgeting, organizational structure, operations, information systems acquisitions and etc. The earlier an agency can foresee and plan for changes the less disruption there will be.

In B.C. the policies and procedures related to the LII, the Land Information Management Framework, have been published but are by no means complete. Ministries can put forth recommendations through the Land Information Strategic Committee, which is a forum for resolution of issues related to land information.

A data sharing mechanism that appears on the surface to reduce human interaction will in all likelihood increase it as agencies work together to make data sharing a reality.

## Conclusions

This paper has reviewed the impacts of data sharing which may cause some initial grief for organizations setting out to increase sharing of information resources. Organizations need to explore more fully those which might adversely affect them, in order that they may be confronted and mitigated.

Extensive data sharing is difficult to implement within a single organization or discipline (Reed, E.O., 1990; Dawson, E.M.J. and Allen, D.E., 1990). Within a diverse collection of connected agencies, as is the B.C. Government, it will not be an simple task. But faced with the need to intensively manage resources which are seen now to be finite, interdependent and vulnerable, decision makers realize that they need the most complete information possible. Combined with fiscal restraint and an increasingly critical public audience, the sharing of data resources is imperative.

## References

British Columbia. Government Land Information Data Exchange Unit. (1991). **Land Information Management Framework**, Victoria.

Dawson, E.M.J. and D.E. Allen (1990). "Data Resource Management at the Surveys and Resource Mapping Branch - B.C. Ministry of Crown Lands", **Annual Conference Proceedings of the Urban and Regional Information Systems Association,** Edmonton, 1990. Vol III.

Durell, W. (1985). "The Politics of Data", **Computerworld**, September 9, 1985.

Edmondson, P.H. (1991). "Addressing Organizational Impacts Affected by the Transition from a Centralized GIS to a Distributed GIS", **Annual Conference Proceedings of the Urban and Regional Information Systems Association**, San Francisco, 1991. Vol II.

Guthrie, J. (1991). "Linking Disparate Data Sources in a Large Organization's GIS", **Annual Conference Proceedings of the Urban and Regional Information Systems Association**, San Francisco, 1991. Vol III.

Hofmeyr H., D.E. Allen and E. Kenk. (1991). "Setting the Stage for Data Sharing in British Columbia", **GIS'91 Proceedings: Applications in a Changing World**, Vancouver, 1991. FRDA Report 153, Forestry Canada.

Kozub, N.E. (1991). "An Exercise in Strategic Planning: Development of Data Policy for the Information Age", **Annual Conference Proceedings of the Urban and Regional Information Systems Association**, San Francisco, 1991. Vol IV.

Leia, G.J. (1989). "Telematics: Canadian Legal Response to Issues of Privacy and Confidentiality", **Annual Conference Proceedings of the Urban and Regional Information Systems Association**, Boston, 1989. Vol. IV.

MacDonald Dettwiler, for the B.C. Ministry of Lands and Parks. (1991). **Land Information Infrastructure: Target System and Migration Strategy**, Vancouver, 1991.

Reed, E. O. (1990). "Using a Distributed Database in Forest Planning", **GIS'90 Proceedings: Making it Work**, Vancouver, 1990. FRDA Report 107, Forestry Canada.

# A3 Corporate data integration

---

**Integrating ecological and forestry data bases using GIS**

Mike Comeau, Environment Canada, Hull, PQ, Canada

**Developing integrated/interfaced geographic information systems**

Yanni Xaio et al., Department of Geography, University of British Columbia, Vancouver, BC, Canada

**Linking a spatially-oriented Northern Spotted Owl population dynamics model with the Western Oregon Digital Database**

Duane Dippon et al., Oregon State Office, Portland, OR, USA

**Multi-ownership integration of forest stand data: A case study in Itasca County, Minnesota**

Greg Arthaud, University of Minnesota, St. Paul, MN, USA

**Alberta's vegetation information system**

Richard Haag et al., Alberta Forestry, Lands and Wildlife, Edmonton, AB, Canada

---

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Integrating ecological and forestry databases using GIS

**M.A.Comeau**

State of the Environment Reporting
Environment Canada
352 St. Joseph Boulevard
Hull, PQ, Canada K1A 0H3

## Abstract

*State of the Environment Reporting makes use of a wide variety of environmental and natural resource variables to assess and monitor the health of ecosystems. Those "working smarter" will recognize that this requires access and linkages to a multitude of existing computerized inventories which must subsequently be integrated and analyzed in a multi-disciplinary manner. GIS and a common spatial framework play an important in this integration process, and this will be the main focus of this paper. The paper deals with the integration of Canada's Forest Inventory, using GIS and a spatial ecosystem framework, with other sources of data maintained or obtained by Environment Canada for State of the Environment Reporting purposes. The Forest Inventory is viewed as an important source of vegetation cover to develop, assess and monitor trends in indicators of ecosystem health. These and other potential uses are evaluated as a result of a pilot area integration project.*

## Introduction

This paper summarizes the methodology and preliminary results of a short-term pilot project conducted by Environmental Information Systems Division (EISD), State of the Environment Reporting, designed to link Environment Canada's Ecological Land Database and Forestry Canada's National Forest Inventory.

The aim of the pilot project was to develop and test a methodology whereby the vegetation component of the Ecological Land Database could be enhanced and updated using the National Forest Inventory. If feasible, and the preliminary results seem to indicate so, this quinquennial source of information could be used as a source for vegetation cover update.

The pilot project is bounded by the Atlantic Provinces but includes those ecological units which cross the Quebec boundary. It is based on the spatial integration, using Geographic Information Systems (GIS) technology, of two distinct but complementary digital map inventories. The primary source of data is Canada's 1986 Forest Inventory, housed within the Petawawa National Forestry Institute, and Environment Canada's Ecological Land Inventory mapped at the ecoregion level.

## Background

In 1979, the former Lands Directorate of Environment Canada initiated the national Ecozones of Canada Project. This resulted in the mapping, description and computerization of some 5456 ecodistrict mapping units. These in turn can and have been generalized for Canada into 177 ecoregions, 45 ecoprovinces and 15 ecozones. The associated attribute data has so far been collected and maintained at the ecodistrict level; the other levels of the hierarchy are used primarily for reporting purposes and as such do not have any associated attributes. The attributes associated with the ecodistrict level include, among other variables, land cover and land activity.

For the pilot project, a regional subset of the ecodistrict database for the Atlantic Provinces, shown in Figure 1, was used.

Forestry Canada has for some time been collecting forestry statistics from forest management agencies. The first computerized national inventory was completed in 1981 and based on some 40 000 standardized map cells. These map cells vary in size and definition depending on the inventory but in general are 100 sq. km. blocks. This inventory was updated in 1986 and will again be updated in 1991. Minor changes in the map cell definitions can be expected with each update cycle.

The 1986 forest inventory statistics are based on sum-maries of 37 individual inventories (14 for the Atlantic Provinces), some up to 40 years old and with an average age of 10 years. While the majority of the derived statistics originate from intensive forest management inventories, other remote areas are based on regional or reconnaissance statistics. While not always available for each inventory, the associated attribute data are stratified by 12 classifiers including data source, ownership, status, land class, site quality, stocking, cause of disturbance, age, maturity, forest type, predominant genus and year of information collection. Each cell has associated with it a number of unique combinations of these classifiers, along with the location code and numerical (acreage) attributes.

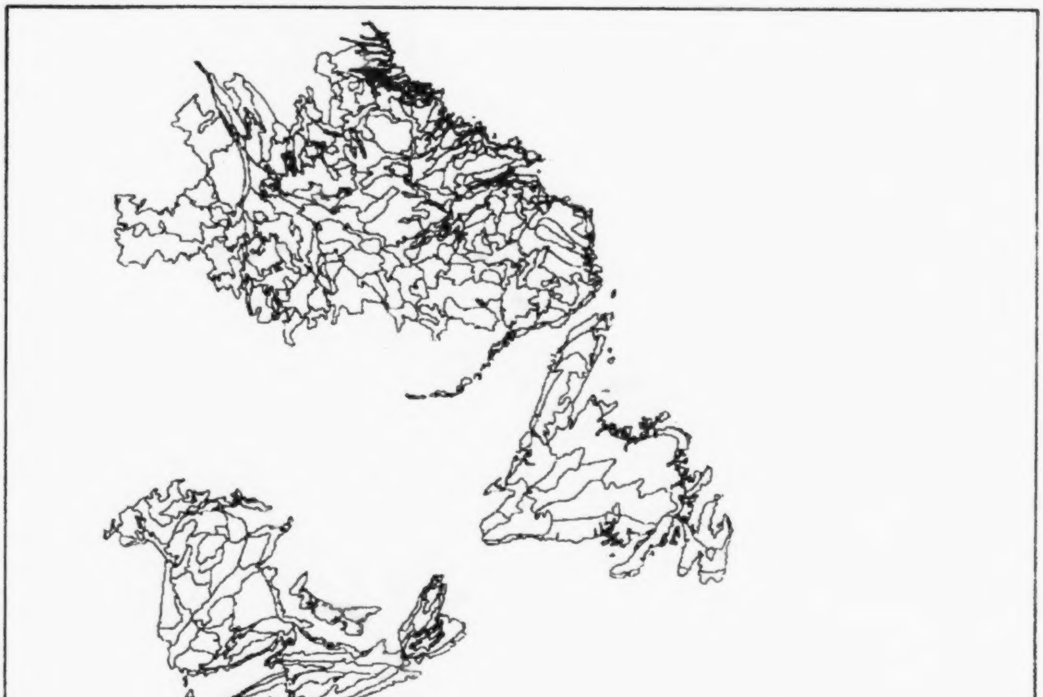
Details of Canada's 1986 Forest Inventory are given in (Figure 1); the forestry cell boundaries for the Atlantic Provinces are shown in Figure 2.

## Methodology

The Atlantic Pilot Project was conducted with the use of the SPANS geographic information system. While this GIS system is different from the ARC/INFO system used by Forestry Canada, industry standard data exchange formats provided the data transfer from one system to the other.

The Forestry grid cells were provided to Environment Canada on a province by province basis. The six provincial grid cell files were vertically integrated into a composite database encompassing all of the Atlantic provinces. They were then horizontally integrated to fit the ecodistrict database and the two

**Figure 1:** Environment Canada's ecological boundaries for the Atlantic provinces



datasets were spatially overlaid. The text that follows describes this integration process in more details and is followed by a preliminary analysis of the overlay results.

## Database loading / vertical integration

A digital version of the Forestry grid were provided to EISD in ARC/EXPORT format. Although vendor specific, this format is similar in contents to the U.S. DLG-30 standard format and, like the DLG-30 standard, has the advantage of being able to support and transfer full polygon topology.

Although provided on a province by province basis, the digital grid files were referenced to a common cartesian coordinate system based on a Lambert Conic Conformal map projection. As is often the case with digital map files, the map projection parameters were not supplied with the exchange format and, in fact, were not clearly known. After consultation with the originators of these digital data (Statistics Canada), and some trial an error, the following map projection parameters were used:

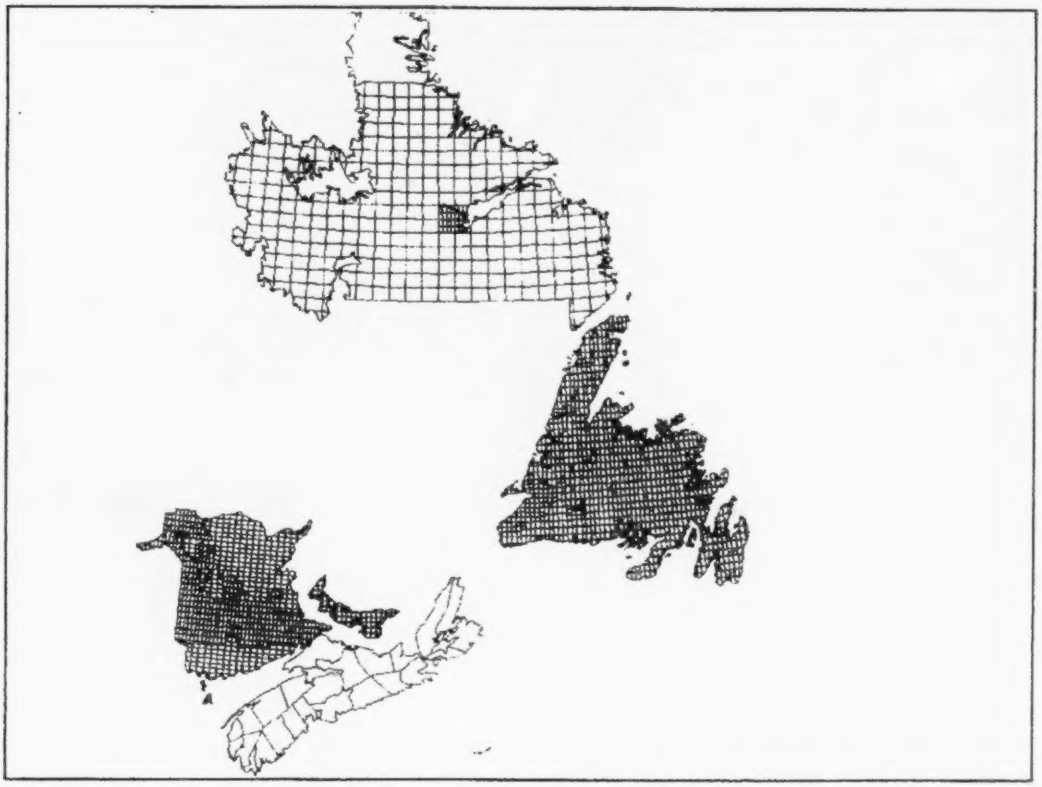
| | |
|---|---|
| Standard Parallels | 49° 10' 00"<br>and<br>77° 50' 00" |
| Longitude of true origin | 91° 50' 00" |
| Latitude of true origin | 63° 23' 26.43" |
| False Easting | 6 200 000 meters |
| False Northing | 3 000 000 meters |

The above parameters seem to give the best overall fit. However some provinces, and especially New Brunswick, deviated in an inconsistent manner from this geographic reference and some local adjustments had to be performed. This was accomplished though "rubber-sheeting" GIS techniques using known geographic reference points on the Forestry grid.

Provincial Forestry files were imported one by one into a SPANSdatabase, transformed to quadtree format, and subsequently merged into one logical dataset. The six files used included Labrador which is considered separately and Quebec which was needed to resolve ecodistricts which span the New Brunswick/Quebec and Labrador/Quebec boundaries. The Quebec cells which were completely outside the study area were eliminated through GIS map overlay and map reclassification techniques.

Unique grid cell identification numbers were maintained as attribute pointers throughout the GIS

**Figure 2:** Forestry Canada's grid cell boundaries for the Atlantic provinces



processing. The corresponding attributes, which included cell identification number, cell area, forest area and unreserved productive forest area, were imported and merged together in dBase. Using the results of previous GIS processing, the attribute records for those cells outside the study area were deleted from the dBase database

A corresponding SPANS compatible ecodistrict database for the Atlantic Provinces was readily available and no details will be given here regarding it's construction. These exist as off-the-shelf items and are a result of some long-term downloading from the mainframe CGIS database. An existing basemap and a map containing provincial boundaries were also utilized to construct GIS overlay masks. Once again the ecodistricts which were outside the study area were eliminated; corresponding attributes were maintained in a dBase database.

The resultant datasets included 2786 Forestry grid cells and 295 unique ecodistricts. The two datasets were topologically overlaid in the SPANS GIS to produce a table and corresponding digital map of all the unique combinations that occur by the superimposition of the two maps. The resultant table, containing 5022 entries, and including a pointer the original forestry cell number, a pointer to the original ecodistrict number, and the area of overlap between the two, was transferred to dBase.

## Basemap registration / horizontal integration

Before elaborating on the statistical analysis that followed the GIS overlay process, some comment on the physical integration of the two datasets is appropriate. This relates more to the contents of the data rather than the format or GIS techniques used to bring the two datasets together under a common geographic reference

Once file formats and geographic referencing have been resolved, there is still the issue of basemap compatibility which has to be considered in this type of data integration process. Basemaps in this context are masks separating land and water which are used to "cookie-cut" thematic map layers. Basemaps are widely used in GIS processing to avoid duplicate digitizing of the shoreline component of thematic layers. In standard GIS practice, the shoreline is digitized once and thematic layers are intentionally digitized outside the shoreline. Sometimes thereafter the thematic layers are masked or "cookie-cut" with the shoreline layer to produce "clean" thematic layers. The dilemma in terms of data integration is at what stage of processing should this masking or cookie-cutting takes place.

In this case the Forestry grid cell layer was already masked with a shoreline layer. Considering differences in map scale, resolution and slight mismatches in geographic reference, one would not expect this thematic layer to fit perfectly with Environment Canada's ecodistrict map layer once it had been masked to a basemap of different scale and different resolution. Indeed there was significant mismatch along the shoreline and the overlay of the two datasets would have resulted in many sliver polygons along the shoreline. These sliver polygons would have either the grid cell or ecodistrict identifiers absent.

Although sliver polygons could have been ignored for purposes of this project, the basemap dilemma was overcome by the fact that the ecodistrict thematic layer was still available in it's uncut form, i.e. with the ecodistrict boundaries extended beyond the shoreline. In the early stages of processing, a study area basemap was constructed using a series of logical map overlays. This new land and water mask was consistent with Forestry's shoreline definition but was spatially extended to include the area where ecodistricts span into the province of Quebec.

Using the SPANS GIS, the ecodistrict map layer did not have to be explicitly masked with this new basemap as this GIS has a unique dynamic basemap capability. The user need only specify which map is to be used as basemap and cookie-cutting is done "on-the-fly".

## Database processing / map overlay

Given the results of the GIS overlay process as a table of unique conditions along with corresponding area, it was then possible to aggregate the Forestry grid cell statistics by ecodistrict using dBase and then re-map such results using the GIS. The ecodistrict identifier made the link between the GIS and dBase and this unique number was maintained throughout the dBase processing.

As a first step, the unique conditions table was sorted by ecodistrict identifier. This table was then related to the Forestry attribute table using a grid cell identifier common to both tables. Using the Forestry grid cell area, acreage of forest area and unreserved productive forest area were converted to percentages and added as two addition columns to the unique conditions table. Using a relatively simple dBase procedure, the weighted sum of both percentage figures was computed as follows:

$$\boxed{\sum \frac{U_i}{E_i} \times P_i}$$

where $U_i$ is the area of the unique condition, $E_i$ is the total ecodistrict area and $P_i$ is the grid cell percentage of forest or productive forest computed earlier.

The above summation was performed for all ecodistricts resulting in a new percentage figure of forested and productive forest area. A subsequent sort and unique index on ecodistrict identifier produced a collapsed table containing a single record for each ecodistrict. This new table could then be related to the

**Figure 3:** Comparison of "forest land" with "forest" land cover.

A. Forest land cover according to Environment Canada

B. Forest land according to Forestry Canada

C. Class comparison on an ecological framework

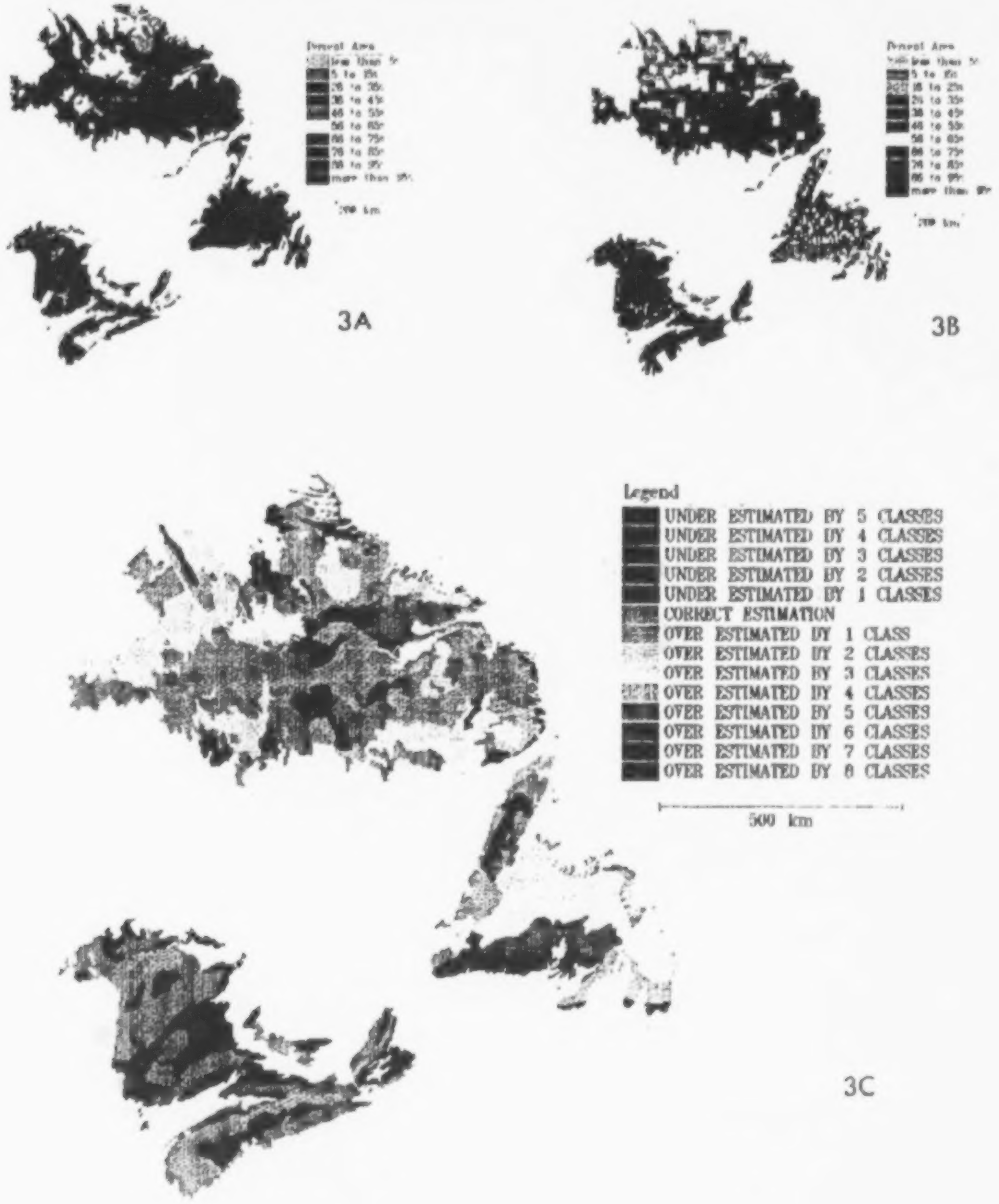

Forest Area
3A
Forest Area
3B
Legend
UNDER ESTIMATED BY 5 CLASSES
UNDER ESTIMATED BY 4 CLASSES
UNDER ESTIMATED BY 3 CLASSES
UNDER ESTIMATED BY 2 CLASSES
UNDER ESTIMATED BY 1 CLASSES
CORRECT ESTIMATION
OVER ESTIMATED BY 1 CLASS
OVER ESTIMATED BY 2 CLASSES
OVER ESTIMATED BY 3 CLASSES
OVER ESTIMATED BY 4 CLASSES
OVER ESTIMATED BY 5 CLASSES
OVER ESTIMATED BY 6 CLASSES
OVER ESTIMATED BY 7 CLASSES
OVER ESTIMATED BY 8 CLASSES
500 km
3C

**Figure 4:** Percentage productive forestry

A. Productive forestry according to Environment Canada

B. Productive forestry according to Forestry Canada

C. Forestry statistics on an ecological framework

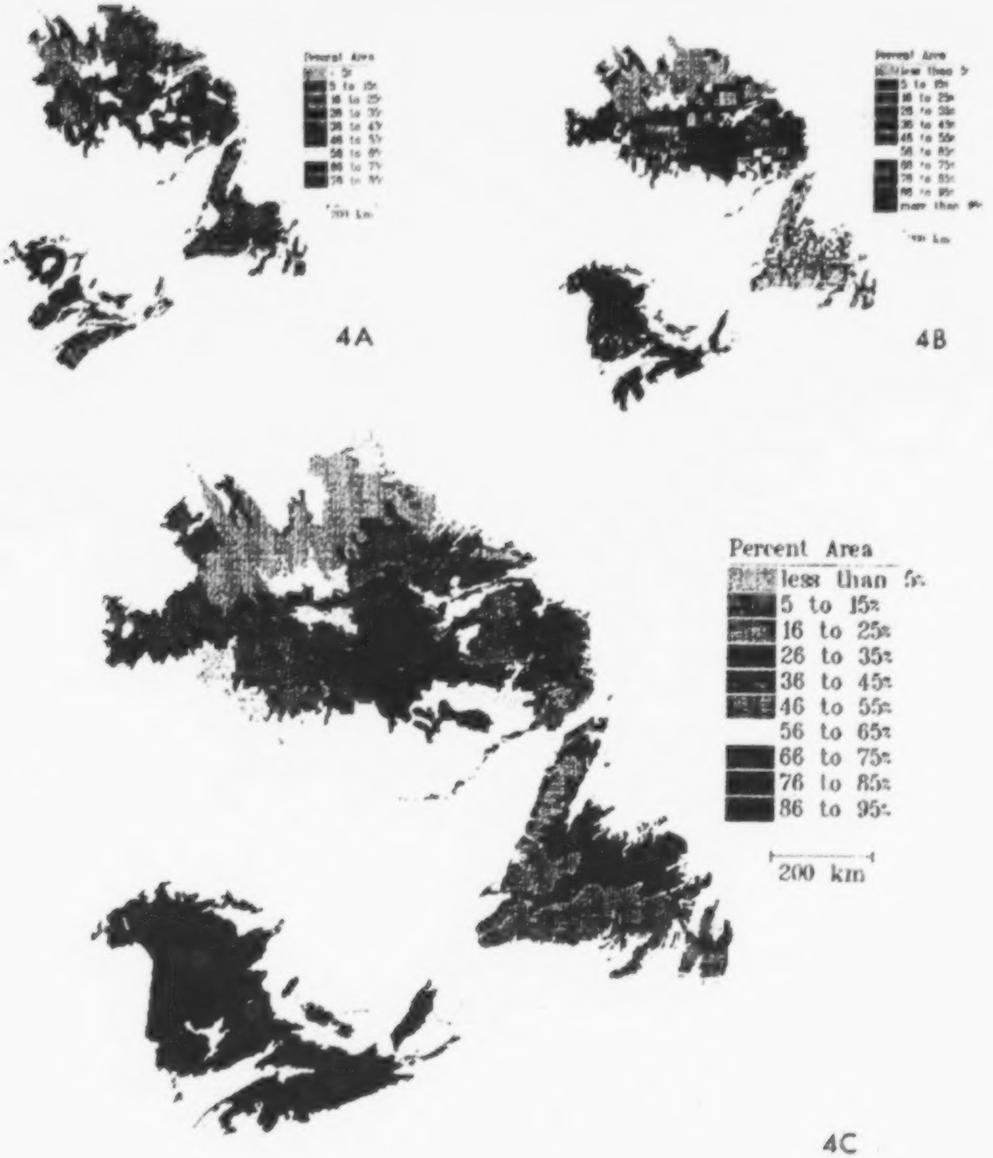
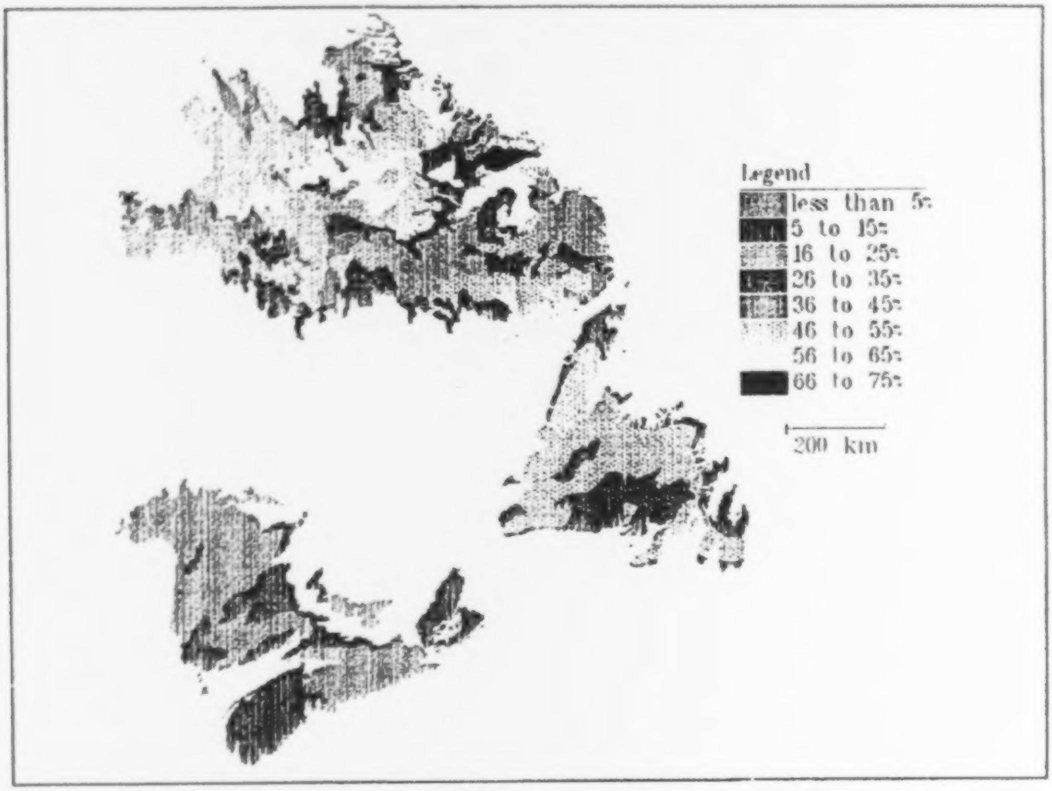

4A
4B
Percent Area
less than 5%
5 to 15%
16 to 25%
26 to 35%
36 to 45%
46 to 55%
56 to 65%
66 to 75%
76 to 85%
86 to 95%
200 km
4C

**Figure 5:** Unproductive forestry
Forestry Canada statistics reported on an ecological framework


Legend
less than 5%
5 to 15%
16 to 25%
26 to 35%
36 to 45%
46 to 55%
56 to 65%
66 to 75%
200 km

original ecodistrict attribute table by unique ecodistrict identifier.

## Preliminary analysis

The land cover component of Environment Canada's ecological land inventory differentiates 12 land cover classes, one of which is "Forest". This class includes wooded lands, including parkland, land boise, open growth forest, forest transition and large woodlots. The value of this class is expressed as a percentage of total land area ranging from 5% to 95% in 10 equal intervals. Forestry on the other hand defines "Forest Land" as land primarily intended for growing, or currently supporting, forest. While their definition includes clearcut lands, a comparison of these two on an equal basis has been made as a first step and the results are presented in map form in Figure 3.

Figure 3A depicts the "Forest" cover class from Environment Canada's Ecological Land Database and is reported here on an ecodistrict basis. Figure 3B depicts the "Forest Land" variable on the forest inventory and is reported on a grid cell framework using the same percentage class intervals used by Environment Canada. The visual comparison of these two does indicate some variance, but it is difficult to quantify the results because of the different spatial frameworks. With the aid of GIS however, a common spatial framework can be applied, and a quantitative comparison of the two datasets appears in Figure 3C.

It is interesting to note some of the regional variations depicted in Figure 3C. The legend is expressed in terms of the class difference between Figure 3A and 3B where a difference of one class represents up to 10% of the area. Based on this comparison, it appears as if the southern coast of Newfoundland and the northern part of Labrador have been grossly over-estimated on the part of Environment Canada in terms of their percentage forest cover. There also appears to be some significant under-estimation on the eastern coast of New Brunswick. The province of Nova Scotia shows some variation as well but the most of these can be explained by nature of the large Forestry grid cells as compared to the smaller ecodistrict units.

If we accept Forestry Canada's definition of "Forest Land", we can then proceed to differentiate between productive and unproductive forest and make comparison with equivalent land activity classes on Environment Canada's inventory. Figure 4a gives the percentage area in productive forestry according to Environment Canada's inventory. Figure 4b represents the same land activity according to Forestry Canada's inventory. With the same GIS techniques used earlier, we can adopt Forestry Canada's figures and simply report them on an ecological framework. The results of this data integration is presented in Figure 4c.

When comparing Figures 4a and 4b, one could conclude that there is an overall under-estimation of the productive forest area on behalf of Environment Canada. In fact, Forestry Canada's inventory has an extra two percentage classes at the top of the scale. Over and above this however we can observe the same over-estimation on the southern coast of Newfoundland and on the northern part of Labrador. It is also interesting to compare Figures 4b and 4c. These are the same statistics; only the reporting unit is different.

As a last exercise in our comparative study we have subtracted the area in productive forest from the total forest area and compared this with the unproductive forestry land activity class in the ecological inventory. Again the same patterns emerge. The final results using Forestry Canada's statistics and Environment Canada's ecological framework appear in Figure 5.

## Summary and conclusions

Based on the results of this pilot project, the integration of Forestry Canada's forest statistics into Environment Canada's Ecological land inventory looks feasible and well worthwhile. Forestry Canada's National Forest Inventory can be used to improve and update Environment Canada's Ecological Land Inventory. This, and other like inventories, could provide continuous baseline data for State of the Environment Reporting without the need for new inventories. The ecological unit framework and GIS are operational tools which will facilitate their integration.

As immediate follow-up to this pilot project, negotiations with provincial agencies are currently underway to obtain other more detailed forest statistics. Their usefulness for State of the Environment Reporting will be further explored. Linkages to the Canada Land Inventory, which gives the land suitability to support forestry, are also being established.

In terms of expanding beyond the Atlantic region, little additional effort other than bulk processing time would be required to complete the same integration in other regions or provinces. It also looks feasible to perform the same integration on a national scale. Although large volumes of Forestry grid cell data will have to be processed, the end result should be easily accommodated in the SPANS GIS.

## References

1 - Gray,S.L. and K. Nietman,1989. Canada's Forest Inventory 1986 - Technical Supplement. Information Report PI-X-86. Petawawa National Forestry Institute, Forestry Canada. 164p.

2 - Wiken,E.B.,1986. Ecozones of Canada. Ecological Land Classification Series, No. 19. Lands Directorate, Environment Canada, Ottawa, Ontario. 26p. and map.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Developing integrated / interfaced geographic information systems

**Yanni Xiao, M.Sc.** (presenting author)
**Brian Klinkenberg, Ph.D.**

Department of Geography
University of British Columbia
Vancouver, BC V6T 1Z2
Tel: (604) 822-6212 / (604) 822-3534

## Abstract

*Increasingly, GIS-related special-purpose programs will be needed for various applications such as location searching and most effective route selection. The users of these programs will usually be people with little or no knowledge of GIS. This situation requires the development of easy to use special-purpose programs seamlessly integrated with the general-purpose raster and / or vector spatial analysis functions which any GIS software is designed to have. Unfortunately, when designing such a special-purpose program you will find that most existing GISs are not designed for this type of use.*

*This paper will focus on the lessons learned from an interface design research project. In particular, one component of the project was the integration of a geographic information system within a public information system shell. The lessons learned from this project should benefit both special-purpose systems developers and GIS developers by highlighting areas where impediments to the design of integrated / interfaced geographic information systems currently exist.*

## Introduction

The issue of public access to government information is a growing concern in the GIS world, mainly because GIS technology can make access to spatial databases much easier than ever before. There are two main concerns about this issue. First, who should be allowed to provide, to the public in a usable form, the information collected by the government? There is a significant push to move public spatial databases, especially at the local level, into the private sector. As a result of this "privatization," citizens or organizations seeking information from government agencies must increasingly rely on data companies. Information disseminated through the private sector is much more remote from the public, however, primarily because it is often prohibitively expensive. Private information vendors, under no obligation to provide the public with low-cost access to government data, often charge exorbitant prices for their services and products. Many public interest groups, therefore, oppose further privatization of public information. The second concern is how to make government information more available to the public? In order to meet the needs of the general public who have little or no knowledge of GIS or other spatial information systems, Public Information Systems (PISs) will need to be developed, by either the government or by alternative sources, in order to make the information accessible to individual users. Only the second concern will be dealt with in this paper.

In the context of this paper, PISs refer to GIS-related special-purpose programs which are designed for some predefined application and for users who are generally not trained in using the system. Some common examples would be location searching and/or route optimization for restaurants, mail or other products pick-up or delivery, emergency vehicle dispatching, airline scheduling, municipal garbage collection, and so on. The basic information in these cases is usually transportation networks. Due to the spatial nature of the information, the development of special software programs interfaced with a GIS—which will provide the general-purpose spatial analysis functions—is thus necessary. The following example presents a typical use of a PIS:

> In a metropolitan area, street networks (the spatial data) can be provided by the agencies responsible for disseminating census data. Address-matching techniques can then be used to add attribute information such as business locations. The spatial and attribute data sets can thus be used together to produce special purpose maps, and to retrieve information for selected geographical areas. Given the name or address of a restaurant, for example, the user may request a street map of the surrounding area and the shortest route to that restaurant from their current location.

There could be many other kinds of Public Information Systems, but their design principles all share similar characteristics. This paper will focus on the network analysis type of PIS only. The conclusions and evaluations presented below reflect experience obtained during the development of a PIS being designed by the BC Telephone Company.

The GIS is expected to provide the following spatial database management and analysis capabilities required of the PIS:

- the display of the base spatial information (e.g., roads, rivers, political boundaries, and other reference features),
- the display of the user queried locations,
- the determination and display of the shortest path between queried locations, and
- the windowing capabilities.

When designing a PIS you will find that most existing GIS systems are not suited for this type of use. In the following two sections of this paper we will discuss the development of a Public Information System and the GIS capabilities required of a PIS, respectively. Although the discussion will focus on the lessons learned while working on the development of an integrated / interfaced geographic information system which was to form one component of a Public Information System, the material is presented is not specific to that project.

## Public information system development

Some decades ago there were few software system installations, and the people interacting with them were primarily computer specialists. Now, we face an increasing number of information systems, the dominance of an interactive work-style, an increasing number of end-users, and new groups of users in various fields [Gimnich, 1990]. In order to make spatial information more available to the public, PIS designs are now in demand. The development of a PIS usually requires the integration of software packages with the following qualities:

- rapid prototyping,
- robust compiled code,
- extensibility,
- a user-friendly windowing (icon-based) environment,
- a GIS as the spatial analysis and spatial data display core,
- linkable to a remote database, and
- potential for multi-media (i.e., audio/visual) display.

In order to achieve these qualities, the developer of the PIS has to consider the following three aspects carefully:

- the platforms on which the PIS will be placed,
- the graphical user interface development facility to be used, and
- the GIS requirements.

Where will the PIS be placed? How will it be used, and by whom? These determine to a large extent the hardware platform that can be used. Considering cost and popularity, a DOS-based PC could be the favored choice for today and the near future. However, when considering the networking and multitasking requirements, the OS/2 (Operating System 2) software platform may be most appropriate. It is believed that DOS and Windows are limited and not powerful or reliable enough to serve as the platform for the new advances in multi-media computing, such as advanced displays, fast computing, powerful video adapters, and special video effects coprocessors, whereas OS/2 is being touted as the multi-media platform of choice. Multi-media is a catch-all term that refers to the integration of enhanced audio and video capabilities into the personal computer. It extends the bandwidth of information that the PC can deliver to the user [Petzold, 1991].

Once the hardware and software platforms are chosen, an appropriate graphical user interface development facility needs to be selected. Fortunately, this kind of facility is already developed and currently available (e.g., EASEL for OS/2 Extended Edition) [IBM, 1990]. Generally, it should have tools which provide developers an accurate WYSIWYG (What You See Is What You Get) environment for creating user interfaces, with features such as primary and secondary windows, pull-down menus, dialog boxes and controls, and icons. Its language statements should be easy to learn and be able to support graphics display, communications, and application logic so that developers can perform numerous tasks [IBM, 1990], such as:

- modify the visual appearance of any object,
- read and write data files,
- reformat and display data graphically,
- provide support for high-resolution color graphics and text,
- interact with dynamic link libraries,
- access relational databases,
- communicate with local and remote applications,
- communicate with host applications.

Since the GIS plays a key role in the development of a spatial data-based PIS, and it is currently the weakest point, we will discuss the GIS requirements separately in the next section.

## GIS requirements

The successful development of a spatial data-based PIS ultimately depends upon the capabilities of the GIS. It is, therefore, very important to clearly identify the characteristics that the ideal system should possess, and then to weigh the cost of using a GIS which lacks some of these ideal characteristics. It is observed that while there are examples of GIS user interface designs that appear to work very well *(in applications for which the interfaces have been specifically designed)* there are many more examples of user interfaces that are designed without serious consideration of the user, particularly when those users are attempting to develop GIS-based PISs.

Generally speaking, a major issue in implementing a GIS is user acceptance. User acceptance is often predicated on the style of interaction permitted by the software. A burgeoning literature focusing on user interface design and development, and on human-computer interaction, indicates that that is currently an area of much active research. This growth reflects the increasing use of computer-based systems, the variety of people who are interacting with them, and the rapid pace of technological advance [Johnson et al., 1989]. However, the problem of user interface software design is a relatively recent concern. User research in computer science is still in its infancy; the term "User Interface Management Systems" (UIMS) was unknown only a few years ago.

With the user's requirements in mind, system designers should consider jobs or tasks when designing the user interface (UI) [Teed, 1991]. The UI is a major component of a total spatial information system user environment. Some practical suggestions and guidelines to aid in the design of the interface between computer systems and their users, developed from expert judgement, common sense, and practical experience, have been proposed in the literature [Brown, 1989]. Brown, for example, points out that a critical step in defining the design philosophy for the user interface is to establish the appropriate balance of ease of learning, ease of use, and functionality. Although there are often trade-offs in designing the user interface, careful attention to human-computer interface design can ensure that features are provided that support all three—we do not have to sacrifice one for another. Some techniques are particularly useful in achieving the balance, the following two are especially relevant to what we have learned from our design project.

*1. Design for novices, experts, and intermittent users.*

Most systems should be designed to incorporate the needs of novices, intermittent users, and experts. Novices may need extensive prompting and rely heavily on menus. Intermittent users may require less prompting, but need to refer to menus occasionally for forgotten details or functions they have not used before. Experts have memorized the sequence of requests they use frequently, and they need not be encumbered by sequences of menus. Experts should be able to shortcut or bypass menus and prompting when desired. They should be able to define complex sequences of actions as a higher order command or macro, such that the entire sequence can be initiated by a single user action. Menus should be available for use by novices, intermittent user, and experts as needed, but time-saving shortcuts should also be available for those who are ready or have to use them.

*2. Provide multiple paths.*

This is perhaps the key for simultaneously providing ease of learning, ease of use, and appropriate functionality. Multiple path techniques including the following:

Menu bypass, which permits the user to request a desired option, display, or transaction directly by entering a comment, code, or other identifier. The user is thus able to bypass a hierarchical menu sequence. Stacking or type-ahead techniques permit the user to make the inputs that would be appropriate for a series of menus or transactions in a single entry. User-defined macros permit the user to save a sequence of individual commands or actions, name the sequence, and later initiate the whole series by entering only the sequence name, code, or function key. Input device options permit the user to initiate a function from more than one input device.

When discussing the characteristics of a GIS the following four subsystems are normally identified: the data input subsystem, the data management subsystem, the data manipulation and analysis subsystem, and the data output or display subsystem.

It is also important, in the context of a PIS, to consider some overall system capabilities.

Using the four main subsystems as a frame, and using knowledge gained from our experience, we can identify some of the requirements that a GIS should have, if it is to form an integral part of a PIS.

*1. Data input*

The data input component converts data from their existing form into one that can be used by the GIS [Aronoff, 1989]. It is important to note that the GIS used for the PIS need not be the GIS that is used by the organization in its day-to-day activities. Also, the spatial information required to meet the needs of the PIS may not form part of the organization's ongoing spatial information requirements. Thus, there may be a high probability that the spatial data will have to be obtained from external sources, such as digital street network data from Statistics Canada (e.g., the Area Master File (AMF) data set). For this reasons the GIS should have extensive import facilities, capable of accepting a wide range of data types. Furthermore, since routing is often one of the analytical requirements of the PIS, it would be preferred if the import facilities maintained as much of the data's intelligence as possible. For example, when importing the road network data from Area Master Files format to GIS internal format, the complete topology of the network should be maintained so that the future network analysis is possible.

*2. Data management*

The data management component of the GIS includes those functions needed to store and retrieve data from the database [Aronoff, 1989]. The data structure and the organization of the database place constraints on the way in which data can be retrieved and the speed of the retrieval operation. The GIS used in a PIS must be capable of rapid data ingestion. As the user interacts with the PIS the results of the user's query must be extracted from the (potentially remote) database, thence passed to the GIS and displayed on the screen within a very short time period. For some GISs this requires the importation of ASCII data, and subsequent translation from the import format to the native spatial data format—all prior to the display of the query locations. This assumes that the coordinate system used by the GIS is compatible with the coordinates stored in the database, and that coordinate translation is not required. This is not always the case, and the developer of the PIS must then weigh the cost of storing additional fields (the GIS internal coordinates) in the database versus performing the coordinate translation on-the-fly.

*3. Data manipulation and analysis*

When using the GIS to perform the network analysis (e.g., to calculate the shortest path between user queried locations), the program should be able to find the shortest path route on-the-fly (e.g., given an initial location and a destination, both of which are subject to change, the shortest path should be calculated quickly and displayed properly). The routing component must be able to display background information in order to provide the PIS user with geo-references (street labels, etc.)—not only the network itself—and ideally the routing software should be able to differentiate between public transportation routes and private transportation routes. The network analysis functions should also allow multiple routing. Given an initial location and several destinations, a shortest path route should be determined which links all destinations without necessarily revisiting a previous site. All these network analysis capabilities are needed for many PISs development, but not all of them are found in most existing GISs.

*4. Data output or display*

For public users of PISs, all of the query results need to be displayed on the screen. Therefore, the quality of display is critical. First of all, it should be very responsive, and the on-screen text (GIS-dependent) should be crisp and clearly legible. The GIS should allow the user to successively "zoom in" to a particular portion of the currently displayed map, or zoom out to provide more of a context. It should not be an 'image' zoom only, and thus limited by the resolution of the initial screen image. Rather, it should be a real map magnification or reduction, capable of increasing or reducing the image resolution as appropriate. In addition, since the PIS will be the primary screen manager, the GIS, therefore, must allow the size and placement of its display window to be fully specified. At best, the window could be redefinable for each individual query; at the least it should allow for redefinitions dependent upon the screen height/width ratio, and the screen resolution (e.g., VGA versus 8514A).

*5. Other overall system requirements*

Some other overall GIS system capabilities are also very important for successful development of PISs. As mentioned above, the GIS used for a PIS must be capable of running entirely in a macro mode in order to allow the PIS designers to bypass the menu hierarchy of the GIS. This is a very basic requirement. Some GISs do possess this feature, but some others do not, or just do so partially. Menus are helpful for novices to the GIS system, but working step-by-step through several levels of a menu hierarchy is slow, and the introduction of a separate (GIS-specific) menuing system would increase the complexity of the PIS beyond that which is commonly acceptable. The GIS used for PIS design should provide support for Dynamic Data Exchange (DDE) (i.e., 'pipes'), and must be capable of running as a background process. DDE is a message-based protocol that allows two applications to establish a connection through which they can exchange data, or commands, in any format that they mutually agree to support. The system must be very robust, and respond to errors without locking up or bombing out. Error handling is a critical feature of effective user interface design of the system [Brown, 1989]. No user error should cause a session to be terminated or aborted, and protection should be provided against costly user errors (i.e., keep user errors from destroying data).

## Conclusions

The dramatically increased number of applications of GISs, such as in the development of spatial data-based public information systems, has placed a heavy demand on their extensibility, and in particular on the design of their user interfaces. Human factors issues and the politics of 'user friendly' are now fashionable topics in professional journals and popular magazines [Shneiderman, 1988]. Since 'knowing the user' is critical to successful system design, those GISs that are developed with careful consideration of every user's requirements (i.e., traditional end-users and value-added users such as PIS developers), and of the general guidelines for human-computer interface design, should prove to be the more successful GISs of the 90's. If GIS designers follow these guidelines the potential benefits to users include reduced training, reduced errors, increased efficiency, and increased satisfaction, while system designers would also benefit through increased user acceptance or sales, improved productivity, and reduced training requirements.

## References

Aronoff, Stan, 1989. Geographic Information Systems: A Management Perspective, WDL Publications, Ottawa, Canada

Brown, C.Marlin, 1989. Human-Computer Interface Design Guidelines. Ablex Publishing Corporation, New Jersey

Gimnich, Rainer, 1990. "A Unifying View on Interaction Styles and Their Implementation in a User Interface Management System", Ergonomics, Vol.33, No.4

IBM, 1990. EASEL for OS/2 Extended Edition — Developer's Guide, Vol.1

Johnson, G.I., C.W. Clegg and S.J. Ravden, 1989. "Towards a Practical Method of User Interface Evaluation," Applied Ergonomics, Vol.20, No.4

Petzold, Charles, 1991. "The Multimedia Extensions for Windows —Enhanced Sound and Video for the PC", Microsoft Systems Journal, Vol.6, No.2

Shneiderman, Ben, 1988. "We Can Design Better User Interfaces: A Review of Human-Computer Interaction Styles", Ergonomics, Vol.31, No.5

Teed, B.E., 1991. "Graphical User Interface: Meeting User Requirements", Spatial Systems Journal, Vol.2, No.2

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Linking a spatially-oriented Northern Spotted Owl population dynamics model with the western Oregon digital database

**Duane Dippon**
**Chris Cadwell**
**Jeff Nighbert**
Oregon State Office
Bureau of Land Management
Portland, Oregon 97208
Kevin McKelvey
USDA Forest Service
Pacific Southwest Expt. Stn.
Arcata, California 95521

## Abstract

*The USDI Bureau of Land Management and the USDA FS PSW Experiment Station have embarked on the development of a population dynamics model for the northern spotted owl. The manner in which the model can be linked to the Bureau's GIS is presented and how the model will be used in the Bureu's planning process is discussed.*

## Introduction

The US Department of the Interior, Bureau of Land Management (BLM administers the use of a wide range of natural resources on nearly 2.4 million acres of federal lands in western Oregon. Two previous papers presented at this Symposium in preceding years (Dippon et al. 1989) and (Dippon and Cadwell, 1991) have explained how the BLM has developed the Western Oregon Digital Database (WODDB) for developing a management plan for the next decade. The earlier paper discussed the process which the BLM is using 90+ resource and base themes to define the operational inventory for each of five common alternatives to be evaluated by the public in the selection of a preferred management plan for the 1990's. The more recent paper explained the manner in which the inventory data was linked to a harvest scheduling model and from the projected harvest level back to sale units on the ground. This paper will focus on the methodology and concepts incorporated to link a spatially oriented, population dynamics model for the northern spotted owl *(Strix occidentalis caurina)* to the current habitat data identified in WODDB, and the future habitat available due to implementation of a management plan for the '90s.

As one component of the BLM planning process, an effort was initiated in 1990 to begin the development of analytical models which would provide measures of the strengths and weaknesses of each of the management alternatives to be examined in this planning process. Just as the estimated ASQ, or Annual Sale Quantity, represents a reflective measure of a plan and it's value to society (employment and revenues), the BLM searched for other models which would help describe the changing health of the managed forest ecosystem.

Several analytical procedures have been under development by teams of resource professionals across western Oregon. The 10-Year Scenario Team has been able to translate the projected decadal harvest to a set of actual sale planning units scattered across the landscape. Teams have also been developing an analytical watershed model and are adapting an elk model. Additionally, the GIS Team has been developing procedures to measure the current level of forest fragmentation across the landscape and how it may change over the next 10 to 100 years. Although these efforts have been interesting in themselves, the owl model represents the tool which the most interest has been focused.

## McKelvey Northern Spotted Owl model

In the Fall of 1990, the BLM funded a project with the USDA FS PSW Experiment Station which would lead to linking the WODDB, and thereby the projected ASQs, with a model which would simulate the changing owl population dynamics across the forest ecosystem as habitat concentrations are altered by harvest removal or in-growth. The project focused on three steps. First, improve on the Lamberson and Noon Owl Model by enhancing it's operability and functionality as a model which can incorporate GIS-based habitat data in a PC-based simulation system. Second, develop a link between the forest inventory data and timber sale units in WODDB which define the appropriateness of individual forest stands for owl survival and reproduction. Third, create a mechanism for projecting the concentrations of habitat across the forest landscape which would result from adopting any of the alternatives.

A spatial model was created to simulate the impact of forest management on populations of the northern spotted owl (McKelvey). The basic premise of the model is that an organism's survival and reproduction can be linked explicitly to its immediate habitat and that habitat's context within the larger landscape. That is, a population's rates of survival and fecundity will vary based on spatial dispersion and continuity. The model allows for both suitable and unsuitable, or marginally suitable, nesting sites to be distributed across the landscape and their placement in relationship to one another will influence the health of the population.

The model simulates each single-organism in the system at any point of time. Each owl is born, moves, attempts to find a mate and breed, and dies. A series of probabilistic rules is used to simulate the behavior of each individual. Individuals birds are born as either male or female and move through the following stages; juveniles (first year birds), sub-adults (second year birds) and adults. Young males look for sites to establish territories while young females look for unpaired males with territories. Pairs split up only when one member dies, or the site becomes unsuitable for nesting (i.e. when logging removes a critical amount of nesting habitat).

In the model, changes in habitat quality are made at the start of any annual cycle, often every 10 years. Habitat suitability affects individual bird survival, fecundity (reproduction success), movement and nesting, or pairing. These parameters can be made to vary as either good or poor, or habitat can be subdivided into as many as six levels, from suitable to marginally suitable, to unsuitable. In landscapes here older forests predominate (>120 years old) or forests less than 80 years old, the proportion of the landscape covered by older forest provides a convenient indicator of habitat quantity (Thomas et al. 1990).

Since any pair can suffer the loss of one member due to normal mortality probabilities, the stability of the population of birds in any given area is dependent on the ability of adjacent pairs to supply the newly formed single with a mate. The more pairs are concentrated in suitable habitat, the more stable the population over time. If the pairs are each in suitable habitat but their sites are scattered across the landscape and are separated by unsuitable, or marginally suitable habitat, then the population will be less stable.

Since the models require knowledge about the suitability of habitat and the movement of birds across the landscape, two units of measure must be defined. What is the scale and time increment? Answering the first helps define the latter. For western Oregon, the model assumes that a pair of birds require 1000 hectares (Thomas et al. 1990). Not all of the home range needs to be suitable habitat, but the higher the proportion of suitable habitat within the home range, the better the survival and fecundity. Based on previous studies (Thomas et al. 1990) indicated that the maximum recorded juvenile dispersal is 100 km. If each home range is identified individually across the landscape as hexagonal cells 1000 ha in size, moving from one cell to another comprises a movement of 3.15 km. The hexagon shape allows the landscape to be fully identified and allows the individual birds to choose from 6 directions to travel in their near-random walks across the landscape as they search for territories (males) or for territorial males (females). Taken together, the data suggests that 32 time increments per year is appropriately. All probabilities for movement, survival and fecundity were proportioned accordingly. However, as with all the parameters used in the model, only by comparing simulation results with empirical data can the parameters be fined tuned to better represent the population dynamics of the northern spotted owl.

## GIS interface

The PC-based simulation model tracks birds across the landscape using a two dimensional matrix of habitat suitability percentages which can be altered over time, providing a third dimension to the analysis. The model in as of itself is not a GIS. However, thematrix requires a GIS to define the spatial distribution of existing and future habitat. The WODDB has several themes which are useful to the model. Forest Inventory (FOI theme) identifies the species, age, crown closure, etc. for every one of the 60,000+ stands which the BLM manages in western Oregon. Since 1986, the BLM has been tagging individual birds and recording their coordinates. All Districts in western Oregon have been spending $100,000's per year since 1989 to tag and determine nesting center coordinates. This data represents another data set identified as the Spotted Owl Centers (SOC) theme.

To conform to the simulation model, both themes must be subdivided into 1000 ha hexagons. For a specific area of suitable habitat surrounded by unsuitable or marginal habitat, the locational placement of the hexagons across the limited space can influence the percent of suitable habitat available. However, the birds probably do not critically eye the environment and determine the optimal placement of home ranges. Besides, if a hexagon theme is identified across the entire region, optimal placement in one location will result suboptimal locations for other locations. With this in mind, the BLM designed a new theme of 1000 ha hexagons. The initial initial hexagon was identified by a point just off the coast at the NW corner of the State Oregon, by Astoria. All of the other hexagons build on the first's coordinates. The theme corresponds with the coverage of WODDB, the Washington State boarder south to the California State boarder and from the Pacific Coast east to the crest of the Cascade Mountains. Every hexagon received a unique subject number which corresponds to it's proper spatial placement in the simulation model matrix.

To relate habitat and initial bird placement to the hexagon grid, the FOI and SOC themes were intersected with the hexagon theme (Figure 1). The resulting data identified which individual forest stands are within any given home range/hexagon. Additionally, the intersection of owl centers provided a record of the manner in which the initial owl pairs and singles were distributed across the landscape. To simplify the analysis, forest stands were assigned to an individual hexagon based on where the majority of the stand resided. With stands averaging less than 20 ha in size, this simplification appeared to instill little or no bias into the modeling procedure. Additionally, where more than one owl centers were identified within the same hexagon, one was arbitrarily moved to and adjacent cell. A simple ascii file can be used to transfer data between the GIS and the owl model.

A key step in the process of linking the GIS to the model was to label each Forest Inventory Unit as to the hexagon it is within. As described in the introduction, the BLM is in the process of evaluating five common alternatives. Each alternative has different land use allocations across the landscape. As described (Dippon et al. 1989), the WODDB and GIS techniques were used to identify the acreage in each forest stand which was assigned to special use, protection, restrictive or intensive forest management. Five databases were created and each was stored in a relational *(see figure 1)* database. By specifying how each stand is linked to the owl model, the current distribution of habitat can be identified for the model's hexagon matrix. Additionally, the future spatial arrangement habitat can be projected for each of the alternatives as well as for the present case.

Once the forest stands were linked to the model, suitable habitat had to identified. The first querery used to identify suitable habitat searched the relational database for all forest stands which had an over story component which was composed of trees averaging 22" in Diameter at Breast Height (DBH), or better, with 40% crown closure; or where the overstory had only a 10% to 39% crown closure but the understory component of the stands was at 12" to 21" in DBH with a 40% crown closure. After review by the Bureau's biologists, it was felt that this stand definition was too restrictive and that it missed identifying some stands which did represent suitable habitat.

A second querery was used to expand from the first. In addition to everything which meet the conditions in the first querery, additional stands were included if their overstory was comprised of trees averaging at least 22" DBH and 10% to 39% crown closure with an understory aged older than 1940 *or* the canopy was comprised of trees averaging at least 12'-21' DBH and a 40% crown closure. Figure 2 demonstrates the varying levels of habitat by hexagon for the two quereries.

After further review by the Bureau's biologists, we came to understand that the forest inventory carries a forester's bias. A good example of such a bias were stands which were generated by fire or heavy partial cutting 50 to 80 years ago. The foresters developed their database to aid in their management of the forest. Since the stand may be composed primarily of 50, 60, or 80 year old trees, that was the information required to plan the proper management of the stand. However, the stand could also have a small component of 200+ year old trees which survived the fire, and windstorms since than. It is the vertical structure of such trees in relationship to the stand as a whole which could make such a stand have suitable habitat qualities.

Therefore, the BLM embarked on an ambitious program to identify which forest stands represents either nesting, roosting/foraging, or unsuitable habitat. Beginning with the stands identified by the two quereries, the biologists reviewed recent photography, stand descriptions and even visiting the sites. Many of the biologists in the Bureau have been in Western Oregon for much of their careers since western Oregon represents the primary forest component of the Bureau's administrative landbase. Additionally, with the intensive tagging program over the last five years, most biologists are very familiar with their landbase.

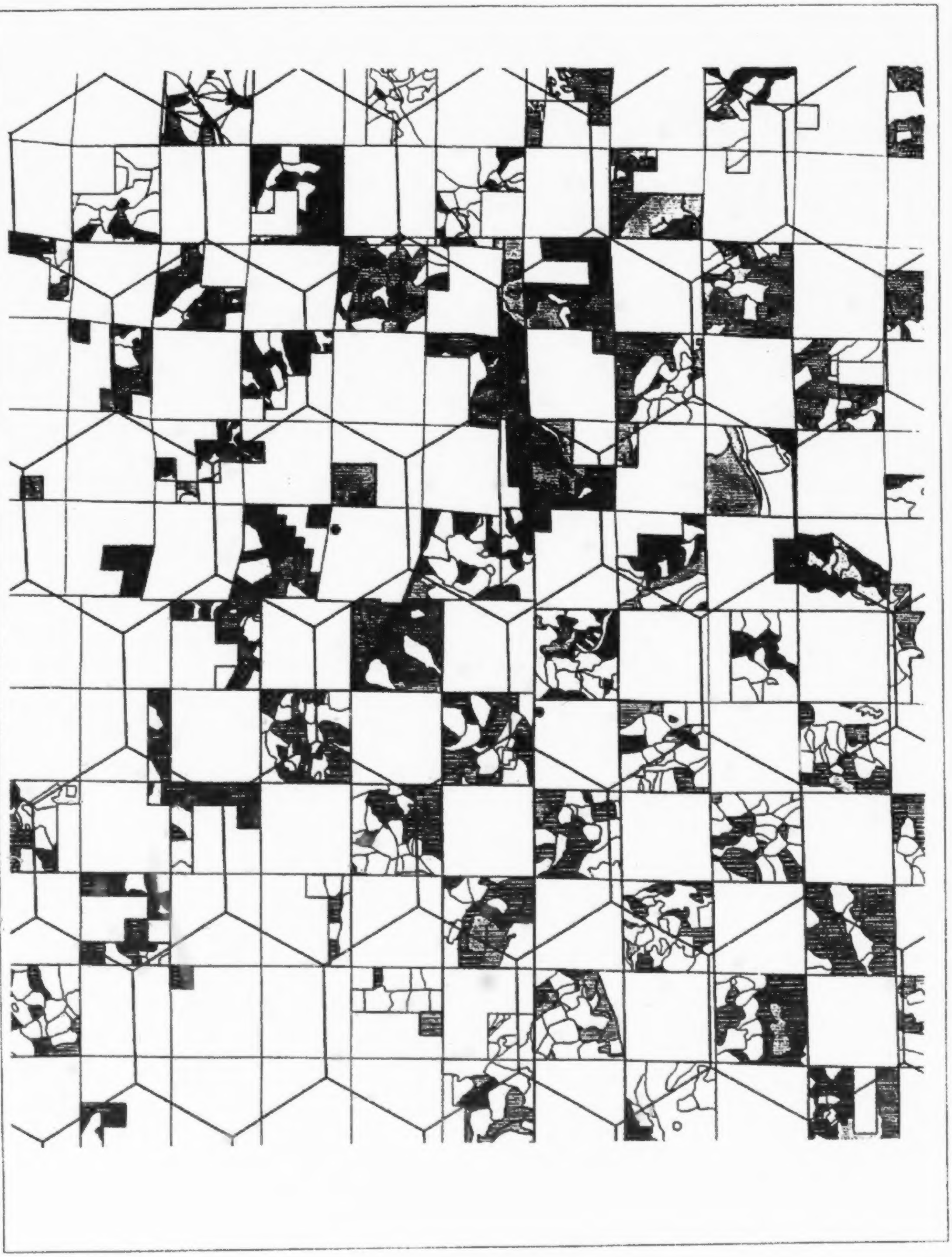


Figure 1 Defining suitable habitat percentages

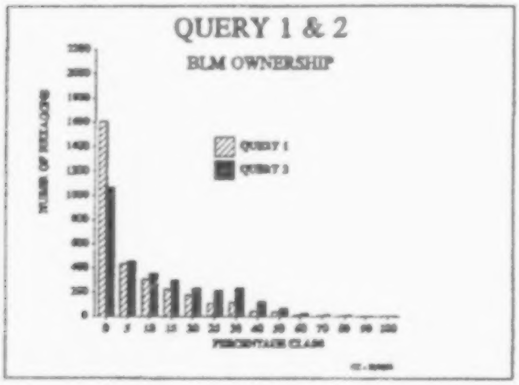


Figure 2

The biologists were told to determine whether an individual forest stand had the "right" structure components which identified nesting or roosting/foraging habitat, based on their knowledge of stands being used for such cases within their resource area. They were instructed to ignore whether there was sufficient quantities of suitable habitat to support the birds, the model would do that. Since individual biologists could apply their knowledge of their area, the definition of suitable habitat varies as the forest ecology varies from the rainy forests in NW Oregon to the much dryer ecosystem in SW Oregon. Differences in elevation could also be recognized. Once the biologists made their selections, a new habitat theme was generated in WODDB by the creation of ascii files of all such tagged forest stand units from the relational database. Figure 2 also presents the result of this endeavor and it shows that more habitat was added with this last effort.

A weakness of the linkage between the WODDB and the model should be noted. The WODDB only has information about the Bureau's land base. As Figure 1 demonstrates, due to the historical allocation of lands in part of the United States, lands administered by the Bureau is intermixed with private ownership. Additionally, the Bureau's lands are normally at lower elevations than the USDA Forest Service's National Forests. Bureau lands share common boundaries with them throughout the State. Since the Bureau's database excludes data from the Forest Service and private ownership, the hexagon suitability percentages are based on BLM only.

The database calculates the number of hectares of nesting and roosting/foraging habitat and divides by 1000 to determine the percentage value for the hexagon. In the case of the Forest Service, their exclusion along the east and west boundaries is a serious problem if the model is to truly project alternative strategies for managing the habitat of the owl. This deficiency is in the process of being rectified. The private ownership lands represent a smaller problem since much of the landbase owned by the private sector has been harvested in the last 80 years and many sites are being harvested for the second time in this century. Assuming that these lands do not currently provide habitat or are not likely to in the future, it is felt that this assumption is an acceptable simplification.

## Analysis of effects

At this stage of the Bureau's planning process, the owl model represents an important component of evaluating the analysis of effects of the alternative management plans under development. As Figure 3 shows, the combined analyses can be used to evaluate the expected population dynamics based on he current disposition of suitable habitat, versus allowing the forests to grow and age over the next 100 years if no harvest is allowed, and versus the long term impact of each of the five common alternatives. Since the spatially oriented land use allocations has been defined within each forest stand, the stand components which are to be managed for timber can be proportionally allocated to suitable or unsuitable habitat categories based on projected silvicultural prescriptions. Most of the stand components allocated to nontimber management status will generally age and become habitat in 100 years.

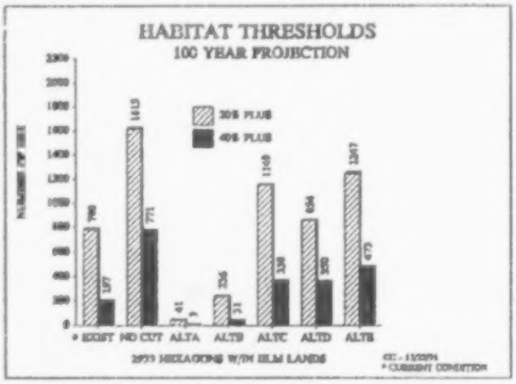


Figure 3

Figure 3 presents the number of hexagons within the WODDB that either achieve 20+ or 40+ percent habitat within the 1000 ha. The difference represents conflicting opinions as to whether 20+ percent suitable habitat versus 40+ percent suitable habitat represents the critical definition as to when a site can be considered "good". While it is important to identify the proportion of hexagons which can be considered "good", it is the spatial distribution and aggregation of these home ranges which have "good" habitat across the landscape, over time, which will determine the viability of the owl population.

Plotting the hexagons which provide 20+ versus 40+ percent suitable habitat provides a useful visual display of the impact that alternative management plans have on the forest ecology throughout the region. Intensive management, represented by Alternative A would eventually place all land into managed stands which would not provide any suitable habitat, only where there currently are major wilderness set-asides. Alternative B represents current management while alternatives C, D, and E represent management plans designed to maintain biodiversity, follow the Interagency Scientific Committee on the Northern Spotted Owl recommendations, or to manage the entire system for non-timber resources; respectively.

Several different forms of output can be generated. The number of pairs over time can be projected. Additionally, an ascii file can be generated to allow the GIS to display the spatial disposition of mean habitat occupancy percentages across the landscape. Such a display will help planners to evaluate the common management alternatives by not only examining the size and stability of the owl population but it will also provide very useful information about the distances and distribution of stable owl home range centers and whether there are a sufficient number when natural catastrophes are considered.

## Summary

The model's analysis incorporating this data is currently being generated. Until the Forest Service's habitat is identified, the model's predictive effectiveness is limited. This problem is expected to be resolved by January. Currently the USDI Fish & Wildlife has a team of scientists developing a recovery plan for the owl throughout it's range in the United States Pacific Northwest. The draft plan is expected to be published in early 1992. The BLM has requested that the two Interior agencies work with the Forest Service to identify the level of habitat for all home range hexagons extending from the US-Canadian boarder to the north, to the north-central part of California, and again, from the Pacific coast to the crest of the Cascade Mountains.

With the northern spotted owl distributed across the region, it is obvious that the health of the species will be dependent of a large federal and state agencies, as well as, on private lands. It is the judgement of the authors that the procedures identified in this paper represent a viable tool for evaluating alternative recovery strategies. While we have only just begun to validate the model, the effort required to develop the suitability percentages in the GIS along with the effort to refine the parameters and hence the behavior of the model, will enhance our collective knowledge on the population dynamics of the owl and the role that the heterogeneous forest ecosystem plays in the owls survival.

## References

Dippon, Duane, Robert Wright and Robert Metzger. 1989. Defining the Operational Forest Inventory for Harvest Planning with a Geographic Information System. In, GIS '89, an International Symposium on GIS. Vancouver, Canada, March 7-10, 1989. pp 55-64.

Dippon, Duane and Chris Cadwell. 1991. Resource Management Planning - Linking a Land Information System to a Harvest Scheduling Model. In, GIS '91, an International Symposium on GIS. Vancouver, Canada, February 12-15, 1991. pp 157-166.

Lamberson, R.H. and B. R. Noon. The Territory Cluster Model: a Spatial Model for the Northern Spotted Owl, Unpublished.

McKelvey, Kevin. A Spatially Explicit Life-History Simulator for the Northern Spotted Owl. Unpublished.

Thomas, J.W., E.D. Forsman. J.B. Lint, B.R. Noon, and J. Verner. 1990. A Conservation Strategy for the Northern Spotted Owl. Report to the Interagency Scientific Committee to address the conservation of the northern spotted owl. Portland. OR.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Multi-ownership integration of forest stand data: A case study in Itasca County, Minnesota

**Greg J. Arthaud**

School of Forestry, University of Georgia,
Athens, Georgia, 30602, USA

## Abstract

*Minnesota is endowed with 16.7 million acres of forest land, 1/3 of the entire state. The U. S. Forest Service manages large areas of federal forest in the Superior and Chippewa National Forests. Minnesota has the second largest state ownership of land in the U.S., much of which is located in our State Forests. Counties in Minnesota own more commercial forest land (2.3 million acres) than in any other state. Additionally, a large portion of forest lands are owned by private corporations, forest industry and other private holders. Not only does there tend to be a lack of consistent management across ownerships, but the data collected on forest stands varies with ownership. Implementation of the project "Impacts of present and potential road systems on timber, tourism, wildlife and aesthetic resources" necessitated production of a seamless geographic information system (GIS) containing stand level data for the project pilot area (two townships). The townships are located in the northern portion of the Chippewa National Forest and ownership (all lands) is 47% federal, 32% state, 4% county and 18% other (mostly private). Problems were encountered with integrating stand spatial data with varying minimum stand size, different stand numbering systems, discrepancies in georeferencing, various definitions of "forest stand", spatial resolution and completeness of data. An all-ownership map was created, digitized and linked to a database designed to provide access to all relevant information within or across ownerships. The completed GIS/DBMS system provides a useful data set to implement spatial modelling and to test the impact of differing ownership management practices on the area.*

## Introduction

*Why the interest in multi-ownership?* Planning across ownerships is of increasing importance. Assume that a parcel of forest land is being considered for management. Adjacent areas, or areas within some proximity can greatly influence the nature of the parcel. Some of the types of activities that might affect the parcel are:

- *Animal movement* Regardless of erected barriers, most animals can traverse ownership boundaries. Planning for protection or enhancement of wildlife habitat requires consideration of characteristics of surrounding areas.
- *Watersheds Activities* performed in a watershed can have profound impacts on downstream sites. In some case, downstream activity can impact upland areas (e.g. dams and salmon spawning).
- *Vehicular traffic* Road systems span all ownerships. Operations taking place in other areas may negatively (e.g. noise) or positively (e.g. increased road maintenance) impact a given ownership area. Joint planning of timber harvests across ownerships could reduce transport costs for the parcel.
- *Forest fires* Fire protection considers all ownerships. Both activities and land

cover type on adjoining ownerships can have large consequences on the parcel.

- *Trail systems* Many trail systems are contained within one ownership (e.g. state park trails). There are many trails that transcend boundaries (e.g. The Appalachian Trail) and should be managed across the ownerships.
- *Timber production* Coordination of activities between owners can increase profitability of timber production and ensure compliance with harvest requirements such as maximum clear-cut sizes. If the maximum clearcut size was twenty hectare with a ten year delay period, cutting of a twenty hectare stand in the current year would impact the ability to cut on the parcel for at least ten years.
- *Insect/disease problems* Insects and diseases do not recognize ownership boundaries. In Minnesota, a major problem has developed with oak wilt. Oak wilt initially infects an oak through a wound, then inflicts most of its damage by travelling through root systems. Construction or harvest activities on adjacent ownerships can threaten oak stands on the parcel. Outbreaks of insects or disease that are not controlled in adjacent areas can impact the parcel.

## Forest ownership in Minnesota

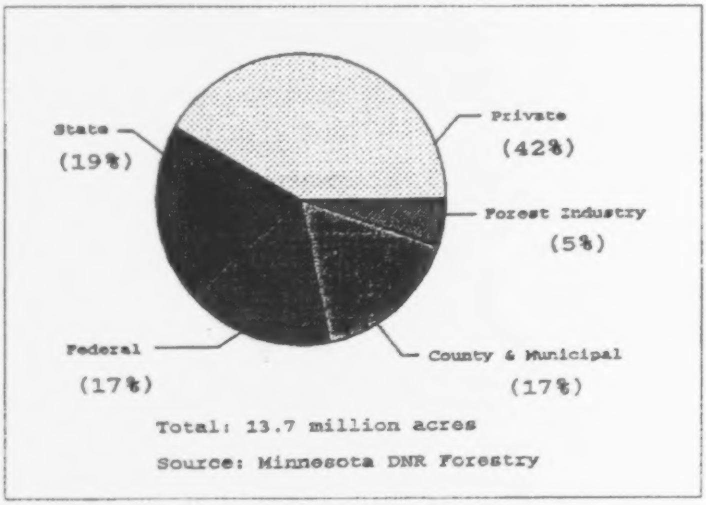
The ownership of commercial forest land in Minnesota, 13.7 million acres, is divided among five major owners (Figure 1). The largest ownership group (42%) of commercial forest land in Minnesota is private, consisting primarily of individual and corporate owners. There is no statewide system to inventory private lands. Many private forest owners rely on the Minnesota Department of Natural Resources (DNR) Private Forest Management (PFM) program or wood using firms for their information.

The State of Minnesota, one of the highest state ownership states, is the second largest ownership group with 19%. State lands have been acquired through federal grants, tax-forfeiture and other means. Large portions of the state land are grouped into state forests. Minnesota has a many State Park and Recreation lands which do not typically allow timber harvesting. State forest inventory consists of a stand inventory (Phase II), operating on approximately a 10 years cycle, and measurement of permanent plots (Phase I). Lands administered by the Division of Forestry are currently being digitized.

Federal ownership of commercial forest land (17%) is primarily located in the Chippewa and Superior National Forests. Stand attribute data is maintained in a computer database and spatial data is updated on individual districts.

County and Municipal forest ownership (17%) is approximately the same as federal ownership. The primary acquisition instrument for counties and municipalities has been through tax-forfeiture. Some counties have used the DNR Phase II system of inventory, while others have developed their own methods, often variations of the state system. Many are digitizing their stand spatial data.

Forest industry controls only 5% of the state commercial forest lands. Much of the industrial land is of high quality and provides stability to industry wood supplies. Access to industrial forest inventories is very limited.



Source: Minnesota DNR Forestry

Figure 1. Ownership of Minnesota's commercial forest land

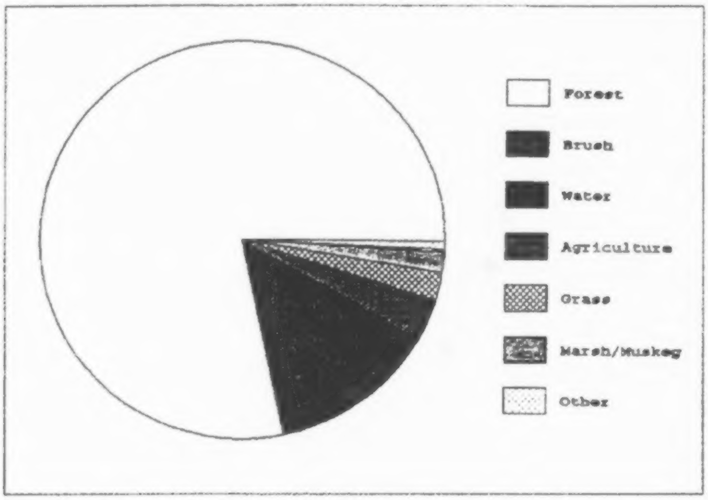


Figure 2. Major land cover types in project area

## The study area

As part of a larger study, "The Impacts of Present and Potential Forest Road Systems on Timber, Tourism, Wildlife, and Aesthetic Resource," forest road data was first collected on several townships in north-central Minnesota. Two of the townships were chosen for the project and are located in Itasca County. The townships are in the northern portion of the Chippewa National Forest and adjacent to the Pine Island State Forest. The area has a rolling topography, with many small and large lowland (brush/grass/marsh/swamp/bog) areas (Figure 2).

The area is used for many activities. Hunting is an important recreation in the townships with several carry-in boat access to area lakes and rivers. Two lakes have boat ramps, primarily used for fishing boats. Trails in the area are used for snowmobiling and hiking. One area, about 100 acres, of virgin pine remains. There are two large management units within the townships; the state administered Dishpan Wildlife Management Area and the Forest Service's Fiske Lake Non-Motorized Recreation Area.

## Ownership in the study area

The U.S. Forest Service is the major forest land owner in the two townships with 7620 hectare (52%). State ownership and administration is the second largest area, 4223 hectares, encompassing 29% of the area. Unlike the state, as a whole, the study area has a smaller proportion of private ownership with only 14% of the total area (2013 hectare). There is a scattering of Itasca County land (550 hectare, 4%).

## What is a stand?

Forest stands are generally defined as a forested geographic area that is distinct from adjacent areas and contains a homogeneous distribution of tree species and sizes. Boundaries of forest stands often are determined subjectively through ground inventory or aerial interpretations. Variables used to manage forest stands, such as cover type, tree diameter and wood volume, may not match those used to photo interpret the area. Smith (1962) defines a stand as "a contiguous group of trees sufficiently uniform in species composition, arrangement of age classes, and condition to be a homogeneous and distinguishable unit." The subjective element of determining what constitutes a forest stand comes from the vague definition of "homogeneity". What constitutes homogeneity for the state ownership may not match that of federal or county ownerships. In inventorying forest stands, consideration should be given to the accuracy of stand variable estimates and the underlying classification schemes used. A stand is best thought of as a biological aggregation of information although, as will be discussed later, management factors often sneak in.

Stand data generally consists of two data sets - spatial data and attribute data. The spatial data is often stored on hard copy maps but is increasing being digitally stored using geographic information systems. To correctly interpret either of these spatial products, procedures used in development should be well documented. Collected stand attribute data varies greatly across ownerships. Often the data is not stored in raw form (e.g. 20 cm diameter tree) but converted to some other aggregated or interpreted form (e.g. tree or stand volume to a given top diameter). To ease integration of data, raw data collection is preferred.

## Federal data

The U.S. Forest Service maintains stand level data for the Chippewa National Forest. Stands are grouped together spatially into compartments, the borders of which are typically rivers, streams and roads. There is no apparent minimum stand size in the data base although "inclusions", areas that are mapped but not present in the database for some reason, are present. Mapped data was available on USGS 7 1/2 minute quadrangle sheets, and was easily digitized.

Since inclusions gave little useful information (there was no attribute data) they were ignored in formation of an integrated map. The Forest Service had implemented several "strip cuts" in the forest and did not distinguish in the database the resulting individual polygons. For this case study, each polygon received a unique label which corresponded to a unique database record. Strip cut data was copied in the database for each of the resulting polygons. Another very interesting feature was concentric circular stands. The U.S.F.S. has created buffer zones surrounding nesting sites and has delineated separate stands around the nests for distances away from the nest. With the current availability of GIS technology, management goals can be stored outside of basic stand data and joined as needed.

## State data

State forest data has been collected for all of the land owned by the state. Digitization of stand spatial data is proceeding and has been completed for the study area. Unfortunately, the available data was not digitized using georeferencing, creating problems with integrating stand maps with other ownership maps. Hard copies of the state maps were created to approximately fit the 7 1/2 minute quad Forest Service maps. Utilizing the associated photo quad's, the state stands were manually added to the Forest Service map and digitized.

As with the Forest Service maps, inclusions were present in the state maps and were not included in the integrated data set. State numbering of stands was by section. The results of numbering based on political boundary were stands with identical stand attribute data, separated spatially by a straight line (section lines). Small stands (less than two hectare) are generally not distinguished in the state inventory system, so important detailed spatial data may not be present.

## County data

The county data inventory was very similar to the state system. Stand were numbered within individual forty's of the section, creating a large number of stands. An area that might be classified a stand using a biological definition could be fragmented into four or more stands using the county system of numbering. County spatial data was stored on hard copy maps with forty's represented as exact square areas. The study area had many sections that could not be classified as "square," resulting in a lack of fit between county mapped data and georeferenced mapped data.

To integrate the county data, forty lines were removed, the stands that had identical attribute data were merged and renumbered and the resulting map was manually added to the U.S.F.S. stand map using the same technique that was used for state data.

## Private data

Private ownership data is a best guess. County soils personnel used aerial interpretation, inspection of adjacent area data and ocular estimate for areas accessible by road to classify private lands. Each stand was estimated for cover type, average diameter and volume per acre. The resulting private stand maps were also not georeferenced, so they were also manually added to the U.S.F.S. maps.

## Putting it all together

Once all ownerships were mapped on the same base map, the U.S.F.S. quadrangle-based stand maps, areas of conflict and areas with no data were investigated. The resulting map included all of the area within the townships and was spatially complete and as geographically correct as possible. A total of 3062 polygons were created, averaging about six hectare per polygon. About 78% of the polygons were forested areas.

Stand attribute data created more difficulties. The U.S.F.S and Minnesota D.N.R. forest inventories contained more detailed data than was available in the county and private databases. Estimates were made for stands lacking important data based first on inspection of adjacent stands with similar data, second on inspection of stands in close proximity to the stand and third on general regression of known stand variables. The result was a unified stand database containing basic information for each stand, such as a unified cover type code, average diameter, total cubic feet per acre, site index (height at age 50), basal area, trees per acre and average stand age.

## Conclusions

The resulting 186 square kilometer stand map, stored in a geographic information system, provides a source for spatial analysis across ownerships. Fortunately, forest stand data was available in some format for the complete area. In most parts of the U.S., large areas of forest stand data is not currently available and would need to be acquired using available techniques (e.g. satellite data or photo interpretation). When the U.S. Forest Service decides on a geographic information system begins compiling spatial data on the Chippewa National Forests, consideration of the problems inherent with their current data sources need to be examined. The activities on non-federal lands within the Chippewa National Forest might have profound affects on management of national forest lands. The need for complete coverage of an area is crucial for modelling of any problem that relies could be influenced by outside area. It could be argued that most modelling problems fit that description.

Other issues not discussed here, but of great importance in integrating forest inventories are the season of inventory (winter is a poor time for capturing regeneration data), the year of inventory, techniques used to project inventory data between stand visits, satellite forest stand estimation procedures, hierarchical forest inventory data structures and inventory design.

It might seem naive at this time to believe that different owners would come together and standardize their inventory techniques, but it should be done. To make future integration of data possible, forest inventories should be designed to capture data in a raw of a format as possible. Inventories that pre-classify tree data or aggregate plot data across the stand are reducing the usefulness of the data that was collected at great cost. Maintenance of individual plot data will allow improved spatial and statistical modelling and may be needed to correctly estimate potential impacts of management, one of the primary reasons for collecting the data.

## Acknowledgements

The author would like to thank Dietmar Rose who acquired the project and the Minnesota Legislative Commission on Minnesota Resources who funded the project. John Rickers and Jonathan Vlaming were key members in the integration process.

## Reference

Smith, David M. (1962) **The Practice of Silviculture**. John Wiley & Sons, New York. 578 pp.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Alberta's vegetation information system

**Richard W. Haag**

Land Information Services Division

**Paul Young**

Automated Information Systems Division
Alberta Forestry, Lands and Wildlife
Resource Information Branch
9945 - 108th Street, 4th Floor
Edmonton, AB
Canada, T5K 2G6
Tel: (403) 427-7222
Fax: (403) 422-0712

## Abstract

*The prevailing climate of fiscal restraint and rising expectations regarding ecological land management make the development of better approaches to handling resource information both a challenge and a necessity. In developing its Vegetation Information System, Alberta Forestry, Lands and Wildlife is focusing on three key factors to meet this challenge. The first is to insure involvement of a broad group of stakeholders. System development was preceded by development of a common set of specifications for a multipurpose vegetation inventory by a Task Force that included both industry and government. In building an environment to manage this inventory, use of a client-server architecture built using prevailing industry standards provides for an open system that will support diverse users. The technical environment will be hinged on a relational database product to insure flexibility, transportability across different hardware and software platforms, and the ability to plan for and respond to rapid change. The second factor is management of risk. The department has been involved in GIS on a project basis since 1985. The development approach capitalizes on this experience, using established techniques with known products. The development plan involves maximizing the use of existing capital resources (hardware, software) through redeployment. The third factor is the focus on integration, both technical and organizational. Development will deliver new GIS capabilities, but it will also support the delivery of data to existing systems to preserve the value of investment in both equipment and data. Development is occurring through a team approach involving resource managers, inventory specialists, GIS specialists and data processing specialists. This assures the integration of GIS into a corporate data environment. Design is being coordinated with the provincial Land-Related Information System (LRIS) Network project and other major departmental initiatives to insure data access and sharing can become a reality. Initial planning has used traditional analysis and modelling techniques to establish scope, define requirements and provide an overall framework for development. Subsequent stages are using a prototyping approach to modular development, beginning with data management and basic product generation components. Initial delivery of decision-support maps and reports that are the result of analysis promises a much faster payback than developments where the initial focus is solely on traditional general-purpose maps.*

## Introduction

Alberta Forestry, Lands and Wildlife is responsible for managing the Crown lands and natural resources under the jurisdiction of its minister. This mandate carries with it responsibilities for a variety of resource management activities including timber harvesting, grazing andindustrial land uses on Crown lands, and management of provincial fish and wildlife populations.

Over the past ten years, several factors have combined to create an increasing demand for natural resource information. Chief among these influences are:

### Continuing development of integrated resource management

The Government of Alberta adopted Integrated Resource Management as a policy in the mid-1970's. Initial emphasis was placed on developing a team approach to land use planning, and on conducting broad-scale Ecological Land Classification to support integrated planning activities. As the integrated approach was assimilated into field operations that had traditionally been sector-based, the definition of "baseline" information for resource management expanded from a base map and specialized resource inventories to include data on topography, hydrography, soils, vegetation, land ownership, access and administration, and the need for more detailed data increased.

Integrated Resource Management not only created a demand for a wider diversity of data. It also required improved abilities to synthesize data, perform a variety of analyses and generate information products to support resource management decisions and allow issues to be presented to the public.

### Changing patterns of resource use

The basic trend has been a shift from reliance on consumption of non-renewable resources to both consumptive and non-consumptive use of renewable ones. Over the past ten years, the proportion of provincial revenues derived from petroleum and natural gas resources, for example, has declined from about three-quarters to slightly less than one-half. More recently, there has been a dramatic increase in the level of forest industry activity in the province.

Increased activity in the forestry sector has created a need to upgrade or replace existing resource inventories with new sources of information that reflect changing resource values (e.g.the increasing importance of deciduous timber) and provide information about other resources (e.g. wildlife habitat, grazing potential). The increase in emphasis on sustainable use of renewable resources has also created a demand for an information base that supports the analysis and modelling of changes over time.

During the same period, however, the general economic climate and a policy of gradual public sector downsizing have combined to produce critical shortages of key natural resource information, and limit the ability to process this information into a form required to support resource management decisions. Priority needs have continued to be met on a project basis, but there is need for a longterm solution to these constraints.

Automation of data and increasing use of GIS technology have shown themselves to have tremendous potential to improve the creation, synthesis, analysis and presentation of information. The challenge has been to devise ways to realize these benefits in an environment of restraint that limits the investment that can be made in new ways of doing business. Our solution has been to build on several key areas of strength, and to adopt an incremental approach that allows new capabilities to be developed on an affordable, modular basis.

## Strengths to build upon

Developing the Vegetation Information System is the first step in implementing a corporate approach to geographic information systems technology. A variety of activities have provided a strong foundation upon which to build for the future.

### Integrated vegetation inventory specifications

Aging of the current, manual forest inventory ("Phase 3") and a need for a greater diversity of information about the vegetation resource were the main driving forces behind development of the Alberta Vegetation Inventory (AVI, 1:20 000 scale), which began in 1987. Initially, specifications for the inventory were developed by a departmental technical committee that included representatives from each of the traditional resource sectors (timber, grazing, wildlife etc.). Initial pilot and prototype work conducted within the department showed that production of a digital inventory was cost-effective compared to the manual inventory it would replace, and that inventory specifications could be expanded to include non-timber values at a reasonable cost. The ability to generate a variety of inventory maps showed the benefits of working in a flexible, digital environment.

Parallel to the work within the department, a joint industry-government Task Force was established that included representatives of Forest Management Agreement holders. This Task Force determined that there was a basic common platform of inventory data that was needed by both industry and government to manage forested lands effectively. This group agreed to test the Alberta Vegetation Inventory specifications that had been developed by the department. The end result was that modifications were made to the original specifications, the Alberta Vegetation Inventory was adopted as a standard by the Task Force, and holders of Forest Management Agreements are producing AVI data.

### An overall strategy

Adopting an incremental approach to GIS development requires a sound framework to insure all the pieces will fit together. Three key elements combine to provide a solid foundation for future development.

The first is a decision made in the late 1970's to begin development of a digital land information base with the production of a new provincial digital base map (1:20 000 scale). This involved resurvey of the province, generation of a common spatial fabric, and the production of both planimetric and digital elevation data. The spatial fabric insures the spatial compatibility of other data sources that will be produced, and the base map itself is a rich source of information on access, hydrography, topography and other key features.

The second is the concept of a hierarchy of systems. At the highest level the provincial Land-Related Information Systems (LRIS) Network, currently under development, is the overarching system that will provide access to "primary" land-related data (e.g. survey control, land ownership/parcel fabric). The LRIS network will also serve as the means of exchange of "thematic" information (e.g. hydrography, access, administration, soils, vegetation, etc.). At a lower level, the Natural Resource Information System (NRIS) is the department's strategic concept for a distributed network of thematic databases and geographic information management tools. It will contribute data to the LRIS Network, and also receive data from it. Building the Vegetation Information System is merely the first step in implementing the departmental NRIS Network, which will in turn be part of the LRIS Network.

The third element is a corporate approach to management of geographic data. When the LRIS and NRIS concepts were first developed, they were seen as something very different from "traditional" computing primarily because of the nature of the data involved. In recent years, however, the value of geographic data as a corporate resource has come to be recognized. Strategic plans for information technology development include geographic data and GIS as part of the corporate information environment, and identify the need to move toward a single network that provides access to all corporate data.

### Considerable experience with computer mapping and GIS

Development of the Alberta Vegetation Inventory began after several years of experience with computer mapping and geographic information system software. Much of this experience took the form of pilot and prototype projects that were undertaken to demonstrate the feasibility of automating the generation of natural resource information products. These projects served to demonstrate the value of automation to key staff members, and they also provided a considerable body of knowledge and experience with a variety of software products and data types. This core of expertise is a key factor in the development of the Vegetation Information System.

## Limiting factors to be overcome

A sound strategic goal, a well-designed inventory, and significant technical expertise will help to make the development of the Vegetation Information System a success, but there are several key limiting factors that need to be addressed.

### The production-development conundrum

Budgets have remained relatively stable in recent years, but escalating demands for information creates the impression of shrinkage. One of the factors that has contributed to this perception is the issue of maintenance vs. enhancement or replacement of existing capabilities. While it may be possible to see great benefits in developing new ways of managing information, it is a fact of life that current levels of service and production need to be maintained. In a "no-growth" environment, this creates a conflict between maintenance and development.

One approach that we are taking to resolve this dilemma is to incorporate the acquisition of hardware for the Vegetation Information System into a general technology replacement program. Most of the equipment that was used to develop the Alberta Vegetation Inventory was originally purchased to support development of the provincial digital base mapping program. As this mini-based equipment is replaced with a workstation-based configuration, the original application-specific systems are being redesigned to reflect a subject-based data driven architecture. With this approach, much of the capital cost of development will be recovered in the form of lower maintenance costs for improved systems. At the same time, existing software is being redeployed to support new applications. This has required an ongoing interaction with both hardware and software vendors, and their willingness to become involved as partners in this transition has been a positive factor.

Another aspect involves a modest, stepwise progression toward the ultimate goal of fully distributed databases and processing capabilities. The starting point is the development of well-designed, subject oriented databases in the form of a central repository, which can be partitioned and distributed in the future. Initial production facilities for map and report generation will also be centralized, and directed at generation of high volumes of standard products.

### Resistance to technological change

It is one thing to develop a strategic view of geographic information as a corporate resource, and another to implement it. There are legitimate concerns about the impact of technological change that need to be addressed: concerns on the part of the traditional information delivery stream (inventory specialists, resource analysts, cartographers) over the role that they will play; concerns on the part of data processing professionals about the impact that a move from a centralized to a distributed approach will have on maintenance of data integrity; and concerns about the integration of data and technology. These concerns are largely a result of the fact that until relatively recently, most of the work done with geographic data and GIS occurred in the head office environment and involved specialists with background in computer mapping and GIS. It did not involve either field staff or traditional data processing professionals to any great extent.

The top-down approach for resolving these concerns has been to establish a Board of Directors for Natural Resource Information. This Board has the responsibility to establish priorities for resolution of information issues, and to insure that approaches to information and technology are integrated. It includes representatives from the operational, resource management divisions, as well as the service divisions that deal with land information and automated information systems.

The bottom-up approach involves the establishment of a development team that includes the same representation. This essentially involves a marriage of two historically distinct cultures: resource management and data processing. Neither of these groups is "in charge". All participants have a role to play, and the use of business, data and process modelling techniques provides a way to resolve different views of the same problems, and to define what future roles should be. Development is occurring at a time of strategic change in the department. All players understand that the last step in any strategic planning initiative is to make structural changes in the organization. Without knowing what these changes might eventually be, each player on the team has a vested interest in the success of every other player. This may not be the neatest solution from an administrative point of view, but it is proving to be a good means to insure synergistic interaction.

### Zealous marketing of technological change

The GIS industry is still in a relatively immature state; a great deal of innovation is occurring, and there is a wide variety of approaches to data structures and software design. "Open Architecture" is a promising idea, but we are still at a stage where many significant software products are produced in different versions for different hardware platforms.

There are many positive aspects to this element of rapid change, including the increase in software functionality, and the rapidly falling costs of hardware. But there is a negative aspect as well, in the form of hesitation on the part of senior management to make substantial investments in what might become "obsolete" next year.

The solution to this situation is to "go with what you know". In developing the Vegetation Information System, we are making use of hardware and software products whose performance has been proven in the pilot-prototype environment. There is recognition that better solutions may be available in the future, but that today's solutions will not become obsolete overnight. We have been fortunate to have good cooperation with vendors in this regard, and consider that the planned configuration involving UNIX workstations, relational software for management of tabular data, commercial GIS software for management of spatial data, and Ethernet for telecommunications will provide a robust solution that will pay for itself and be adaptable to future changes.

## The development path

The path that we have adopted for developing and implementing the Vegetation Information System recognizes these strengths and limiting factors, and takes into account the current limitation on telecommunications capacity and the state of technology for managing distributed databases. It focuses on delivering a series of centrally-configured functional modules which can be distributed as technology develops and financial resources become available. Four modules are planned.

### Data Storage and Maintenance

The current stock of Alberta Vegetation Inventory data exists in the form of digital map files and "coverages" containing polygon data, and a variety of files containing point (particularly volume sampling) data. One of the key steps in building this module involves a restructuring of the inventory from its current map-based format to a normalized data structure, based on entity-relationship modelling. This will provide a much greater degree of flexibility in applications development (e.g. selection of stands that meet certain criteria and calculation of associated timber volumes), and it will make it possible to accommodate future changes in inventory specifications.

Restructuring is also necessary in order to accommodate the storage of inventory data being produced by holders of Forest Management Agreements. A normalized data structure will make it possible to treat submissions of new inventory as transactions, so that concerns about confidentiality and distribution of data on the part of suppliers can be respected.

One of the things that data modelling has shown is that, in order to be useful to most resource managers, vegetation inventory data has to be available in combination with other data types, particularly administrative boundaries. The solution has been to include this information in the database design, and to provide for its production through existing mapping programs.

The physical design of this module includes the use of relational database software for management of attribute data, and the use of commercial GIS software (ARC/INFO) for the management of spatial data. Although the strategic goal is a common structure for all forms of data, this hybrid approach provides a workable solution for the medium term.

The hardware platform for data storage and maintenance will consist of a DECStation 5000, connected to the departmental Ethernet network.

### Basic product generation

The initial priority for generating information products from AVI data was development of a standard vegetation map product to replace the current Phase 3 Forest Inventory maps used by the department as the primary source of vegetation information. This priority reflected a need to demonstrate to operational resource managers that digital technology could fit into and support departmental operations, rather than forcing operations to adapt to it. It was also a result of the fact that production staff were more familiar and comfortable with computer assisted mapping than they were with GIS technology, and felt this was the best use of limited production capacity.

Prototype work showed that it was feasible to produce the equivalent of map products that had been produced from the previous generation of manual inventory, essentially summaries of inventory data. Extensive review has shown that there is considerable demand for these "information summary" products, and they will be part of the suite of products delivered by this module.

The Basic Product Generation module will also deliver two other kinds of products: "custom maps" and statistical reports. The term "custom maps" refers to products that are the result of analysis (usually reclassification) designed to support a specific resource management decision. A good example is a series of merchantability maps that have been developed for different forested regions of the province. The identification of merchantable stands is one of the first steps in harvest planning, and the use of custom products has shown itself to reduce the effort required to identify these stands by about 80% compared to manual methods.

Statistical reports on, e.g., stand areas and timber volumes are essential for determining allowable harvest levels. Conversion of the map-based inventory to a normalized structure, provision of administrative boundary data and the use of relational database management software will make it possible to generate these reports efficiently. These reports are currently generated from the Phase 3 Inventory on a mainframe system which is expensive to maintain and modify. As Phase 3 data is either converted to the AVI data structure where inventory quality and longevity warrant, or replaced by AVI data, this reporting function can be assumed by the Vegetation Information System. This will have the effect of increasing product diversity and flexibility, and reducing maintenance costs.

The initial configuration for this module will be centralized, due to the limited availability of hardware in regional locations and limited telecommunications capacity. High volume production will be achieved using a color electrostatic plotter that will be shared with other mapping programs and funded on a cost-recovery basis. Product generation can be regionalized over the short term using GIS software on either microcomputers or workstations, to the extent that budgets permit. For the next several years, however, regional product generation will depend on distribution of data via hard media.

### Update

The current mechanism for updating the Phase 3 Forest Inventory is based on interpreting aerial photography, compiling changes, redrafting existing maps and manually updating mainframe attribute files. Update is performed to identify changes in forest cover due to harvesting, fire, depletions due to land use, and insect infestations. There is a nominal 5 year cycle, but limitations on manpower and availability of aerial photography often extend this cycle.

The goals for this module include reducing data capture costs, reducing the length of the update cycle, and providing a means to detect and analyze changes to the inventory over time. Development is currently in a pilot phase in which the potential to use remotely-sensed (satellite) data is being assessed. Abilities to detect changes, locate them accurately and classify them properly are issues that will be resolved.

Regardless of whether satellite or airphoto data are used to update the inventory, this module will be implemented initially in a central configuration. Database design will accommodate online storage of historical data, in order to support analysis of changes over time.

The update module is being designed to detect and classify changes in vegetation cover due to timber harvesting, land use activities and natural disasters. There is another aspect of change over time that is important to the resource managers who will use the Vegetation Information System: gradual changes that are the result of ecological processes including growth and succession. Modelling and analyzing these changes is one of the functions of the fourth module.

### Decision-support applications

Building the Storage, Basic Product Generation and Update modules will provide an infrastructure to support departmental and government-wide needs for information about vegetation. The Decision-Support Applications module will have a different purpose: supporting particular business needs that have a requirement for sophisticated analytic modelling capabilities and/or specialized data.

This module will consist of a variety of applications, e.g. timber supply analysis, wildlife habitat modelling and integrated resource planning. Development of each of these applications will make use of the same team of information and technology specialists as other modules. It will also require a sponsor to provide specialized business expertise and undertake responsibility for the additional funding that will be required.

Involvement of both government and industry will continue as these applications are defined and developed. This module is currently in a preliminary planning stage, and possibilities are being pursued actively through discussions with industry and through the establishment of provincial GIS user groups that provide a forum for exchange of ideas about needs and possibilities.

## Conclusion

Developing the Vegetation Information System is the first step in implementing a corporate approach to geographic information systems in Alberta Forestry, Lands and Wildlife. The starting point is increasing the availability of data, including vegetation inventory and other key sources of information. The initial modules will provide standard map products for those who do not have the capability to handle digital data. It will also provide custom products that will reduce the time required to compile and analyze information and allow resource managers to direct their attention to the management of resources rather than data. The combination of product delivery and an open approach to technical architecture promises to spread benefits across the organization, providing rapid payback and a flexible path toward future growth.

# B1 Inventory and planning

**Approaches to forest survey, inventory, evaluation and planning using biophysical models of landscape processes**

T.W. Norton et al., The Australian National University, Canberra, ACT, Australia

**Forest monitoring using temporal GIS techniques**

Gail Langran, Intergraph Corporation, Seattle, WA, USA

**Analysis of the spatial distribution of a forest population for the planning of a forest inventory**

Jean-Marie Bilodeau et al., Université Laval, Sainte Foy, PQ, Canada

**Quality control in data base development and spatial analysis**

Brendan Mackay et al., Forestry Canada — Ontario Region, Sault St. Marie, ON, Canada

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Approaches to forest survey, inventory, evaluation and planning using biophysical models of landscape processes

**T.W. Norton**
Research Fellow*

**H.M. Neave**
Doctoral candidate*

**H.A. Nix**
Professor and Centre Director*

**I.D. Moore**
Jack Beale Chair of Water Research*

*Centre for Resource and Environmental Studies
The Australian National University, GPO Box 4
Canberra, Australia 2601

## Abstract

*Integrated forest management demands greater consideration and knowledge of the dynamics of many complex biophysical patterns and processes in forests. In the past, realising these needs has been hampered by many factors including the ability to manipulate large data sets so as to realistically and efficiently model complex systems in three-dimensional landscapes. This paper briefly considers recent progress towards integrating heirarchial analysis, digital terrain analysis, bioclimatic modelling, hydrological modelling and biological modelling into a GIS-based framework permitting more quantitative and dynamic analyses of dominant landscape patterns and processes. Some applications of the general framework and modelling approaches are then discussed, including the design of a field study to sample the regional variation in species richness and abundance of forest avifauna.*

## Introduction

The world's native forests support many indigenous peoples and are used for a variety of purposes including as watersheds, for nature conservation, for spiritual enlightenment and recreation, and for wood production. Currently, the way forests are allocated and managed for many of these uses is controversial. While it now appears generally accepted that forest usage should be sustainable, exactly what this means in the context of different forests, different uses and for different intensities of use in space and time is often unclear. What is clear, however, is that if forests are to be allocated and managed on an ecologically-sustainable basis, then one important prerequisite will be the availability of much better techniques than are currently employed routinely to help survey, inventory, evaluate, plan and manage.

In this paper we briefly consider some approaches to forest survey, inventory, evaluation and planning using biophysical models of landscape processes as a basis. Specifically, we discuss progress towards integrating heirarchial analysis, digital terrain analysis, bioclimatic modelling, hydrological modelling and biological modelling into a GIS-based framework to permit more quantitative and dynamic analyses of dominant patterns and processes within forest environs. Some applications of these approaches are then considered including the design of a field study to sample the regional variation in species richness and abundance of forest avifauna.

The research reported here has been conducted primarily in Australia and trialled largely using Australian case studies. However, the techniques and framework are generic and, more recently, are being employed increasingly in a variety of terrestrial landscapes on other continents (see, eg. CRES 1991, Moore *et al.* 1991, B.G. Mackey, Visiting Research Scientist, Forestry Canada - Ontario Region, pers. comm.).

## Operational framework

The generalised, operational framework for modelling landscape pattern and process which underpins much of this paper's discussion was articulated by Nix and Gillison (1985) and, more recently, by Mackey and Bayes (1989) and Norton *et al.* (1990). Although Nix and Gillison (1985) specifically addressed forest habitat and wildlife management, the conceptualisation was generic in that it could be applied to most aspects of land inventory, survey, evaluation and planning. The argument was, in effect, that the sustainable use of forests rested largely on an understanding of process. The authors described methodologies that aimed at more quantitative and functional definitions of the habitat and environmental "envelope" of forest taxa which could be employed for predicting their distribution and abundance and, ultimately, the likely impacts of different management practices on wildlife. Predictions were generated with user-specified models coupled to a spatially-referenced data base (GIS) consisting of those biophysical attributes that regulate the primary radiation, thermal, moisture and nutrient regimes. The minimum data set for modelling comprised only those climate, terrain and geological attributes necessary to estimate these regimes (Nix and Gillison 1985).

The operational framework consists of a model of the proposed use (eg. to characterise regional variation in certain climatic attributes, to predict the distribution and abundance of wildlife) that incorporates the biophysical and other factors known to most strongly influence the model's performance. The model is coupled to a GIS data base that contains the biophysical attributes specified in the model and a prediction is generated for a given area. At this point, predictions must be tested, and the model refined before it can be used to formulate management prescriptions. Alternatively, if the model's performance is unacceptably poor, the model may be discarded.

Nix and Gillison (1985) stressed that complete understanding of a system was not required to derive useful predictions. They argued that provided the major processes are considered in modelling, relatively simple and robust models can have predictive value. One additional and significant advantage of the proposed framework was that it provided considerable flexibility and allowed for refinement and innovation as new knowledge evolved. This was in contrast to many of the prevailing methods of habitat definition which locked future research and development into a single, fixed *a priori* system of forest classification. The latter limitation still persists in many circumstances (Norton *et al.* 1990, Norton *et al.* 1992).

## Minimum site-attribute data sets

A critical component of the operational framework was the ability to (i) derive reliable estimates of the minimum site-attribute data (climate, terrain, substrate) for large regions and at relatively fine scales; and (ii) to be able to manipulate and analyse these data efficiently and quickly. The latter need was met, obviously, by the advent of modern computers. The former need was met by the development of a range of mathematical techniques. Both areas are under continual refinement.

In the case of climate, Hutchinson (1987) has developed and successfully applied spatial interpolation techniques to long-term meteorological data to generate estimates of a variety of long-term, mean (monthly, weekly, daily) climate values for any point in Australia for which latitude, longitude and elevation are known. When these models are coupled to a digital elevation model (DEM - regular grid of elevations approximating landscape form or terrain; see below), estimates of climate can be derived following a regular grid. Thus, for example, mean daily maximum and minimum temperatures across Australia can be described by month with standard errors not exceeding about 1 degree Celcius, monthly mean solar radiation with errors not exceeding five per cent and monthly mean precipitation with errors not exceeding about 15% but usually around five per cent. In other words, the error associated with these estimated values is typically of the same order as that associated with observer and instrument error.

Elevation can now be represented by digital elevation models. From these, terrain models of slope and aspect and catchment-area calculations can be derived (Moore *et al.* 1991). A technique for interpolating accurate, continent-wide DEMs at a suitable level of generalisation from minimal-surface, specific-point elevational and streamline data has been developed by Hutchinson (1989). This technique has been used to generate a rectangular-grid DEM for Australia with a horizontal resolution of 1/40 degree of latitude and longitude (approximately 2.5 km) (Hutchinson and Dowling 1991). Using a similar technique, DEMs have also been created for several regions of the Australian continent to undertake various survey, inventory and evaluation studies (Mackey *et al.* 1989, Norton *et al.* 1992). These DEMs have horizontal resolutions as low as about 9 seconds (approximately 200 m).

Currently, estimating spatial variation in soil nutrient regimes is limited since compatible soils data are not available at useful spatial resolutions for most of the continent. Geological substrate is mapped at a scale of 1: 250,000 for the entire continent and represents the best available data set for the derivation of mineral nutrient regimes. Mackey *et al.* (1988, 1989) reported a procedure to derive a spatial estimate of soil nutrients for relatively large regions by assigning a potential nutrient supply index rating (0 - 10) for the substrate to each lithological unit that had been digitised from geological maps of the region. The assignment of values was done in close consultation with a group of soils experts (Mackey *et al.* 1988). This method has since been adopted in the study of Richards *et al.* (1990) who developed a spatially-referenced set of biophysical data for the south east forest region of the State of New South Wales as a precursor for assessing aspects of its biological conservation. This region is over one million ha in area.

## Complementary site-attribute data

Several additional site-attribute data sets such as extant forest cover, wildlife distributional data and data on the history of site disturbance are important, if not essential, for many forest survey, inventory, evaluation and planning tasks (Norton *et al.* 1990). The present distribution of forest canopy, for example, can be obtained in digital form from the satellite-borne LANDSAT MSS sensor. Since a number of difficulties arise in using LANDSAT data to identify floristic and structural forest types, the most useful application of these data to date has been in simply indicating which areas retain or do not retain forest cover (Norton and Nix 1991, Norton *et al. 1992).*

Museums, herbaria and a range of other research agencies represent a valuable store of locational data for forest taxa in Australia. These data may be used to help develop predictive models of wildlife distributions or to help devise wildlife surveys (Nix 1986, Davey and Norton 1990, Busby 1991, Nix and Switzer 1991).

## Development, refinement and applications

In the last several years the general operational framework has been under active development and testing. Techniques such as bioclimatic modelling using the Bioclimatic Prediction System (BIOCLIM - Nix 1986, Lindenmayer *et al.* 1991, Norton *et al.* 1992) have been extended and employed in a heirarchial manner with terrain, hydrological and biological models to undertake different modelling tasks (Richards *et al.* 1990, Moore *et al.* 1991).

For example, to identify potential habitat for the Koala, *Phascolarctos cinereus* in south east Australia, Norton and Moore (1991) used BIOCLIM, terrain, substrate and forest cover data held at a relatively coarse resolution to locate forested water catchments most likely to support habitat. Within those areas, much more detailed terrain and hydrological modelling studies (eg. upslope catchment index, soil wetness index) were undertaken using grid-based analyses at a scale of a few tens of metres to 'fine-tune' the spatial and temporal predictions. The latter modelling techniques were originally developed for engineering and soil erosion evaluation (Moore *et al.* 1991).

Many environmental tasks have been addressed using these types of modelling approaches, at a range of scales. These tasks include:

**(1) to characterise the spatio-temporal variation in physical or biophysical attributes of landscapes as a basis for:**

- systematically stratifying the location of sites for forest survey and inventory work (Austin and Heyligers 1989, Neave *et al.* 1991);
- the systematic collection of out-standing site-based data for building predictive models (Norton and Williams 1990, Lindenmayer *et al.* 1991, Mackey 1991);
- assessing the representativeness of sites for nature conservation (Mackey *et al.* 1988, 1989, Margules and Stein 1989);
- quantifying the realised niche of tree species and/or provenances to help the identification of new sites best suited to grow certain species/genotypes or to identify the best trees/genotypes to grow at a particular site (Booth *et al.* 1988);
- assessing the potential viability of existing wildlife corridors or the potential to create new corridors (Norton and Nix 1991);

**(2) using scientific understanding and models of relationships between biophysical pattern and process to:**

- predict the distribution and abundance (ie. habitat quality) of taxa in space and time (Lindenmayer *et al.* 1991, Norton *et al.* 1992);
- predict spatial variation in forest architecture and physiognomy (Mackey and Bayes 1989, Mackey 1991, Bayes and Mackey 1991);
- predict variation in site productivity and the rate of growth of trees; and

**(3) to evaluate nature conservation and reserve design strategies for large regions:**

- eg. Mackey *et al.* (1989) in north east Queensland tropical rainforest and Richards *et al.* (1990) in south east New South Wales temperate forests.

## Case study: regional variation in species richness of birds

To illustrate the utility of a process-based modelling framework consider the design of a field study to sample the regional variation in species richness and abundance of diurnal forest birds (ie. one application of point 1, above). Two questions common to many field studies were addressed: where to sample, and how to use site-based observations for prediction beyond the site (ie. spatial extension). The target region for this study was the south east corner of Australia encompassing S.E. New South Wales and East Gippsland, which represents an area in excess of 1 million ha.

By using a GIS containing climate, terrain, substrate and forest cover data stored as a grid with a 9 second horizontal resolution, it was possible to characterise some important components of the meso-scale environmental variation within existing forests into 20 climatic categories on soils of relatively low and high fertility. These categories were used as the first stratification in the selection of field sites for sampling birds. Further stratifications were slope (north facing, south facing) and topographic position (ridge, mid slope, lower slope) which can be used, *inter alia*, to estimate solar radiation and water balance regimes. Allowing for the need to replicate sites with each combination of these attributes to develop modelling capability, a total of 164 sites were identified for sampling across the region. For bird sampling, a one ha plot size was chosen. Thus, the number of sites identified by this approach to span various regional environmental continua represents a very small percentage of the total area. As data on the species richness and abundance of birds versus site-attributes are collected, it should be possible to develop quantitative models to predict the regional variation in these and other parameters of bird communities.

## Conclusion

The research described above was borne out of a clear need for more reliable and efficiently collected information to improve decision-making about the allocation and management of landscapes. Computer-based, GIS technology is revolutionising how forest survey, inventory, evaluation and planning tasks are undertaken and is clearly the most efficient way to manipulate and analyse large resource data sets. These technologies help create a dynamic work environment more akin to the dynamic nature of the forests that we attempt to characterise, understand and manage. However, a move from a static to a dynamic or process-based analytical and predictive capacity cannot be resolved by such technology alone. Rather, it is essential to explore, develop, refine and apply research paradigms and operational frameworks concerned with understanding and modelling the linkages between pattern and process.

## Acknowledgements

Helen Neave is funded by an Australian Postgraduate Research Award from the Federal Department of Employment, Education and Training. Professor Ian D. Moore thanks the Water Research Foundation of Australia Ltd. for assistance.

## References

Austin, M.P. and P.C. Heyligers (1989). Vegetation survey design for conservation: gradsect sampling of forests in north-eastern New South Wales. **Biol. Conserv. 50: 13-32.**

Bayes, A. and B.G. Mackey (1991). Algorithms for monotonic functions and their application to ecological studies in vegetation science. **Ecol. Modell.** 56: 135-159.

Booth, T.H., Nix, H.A., Hutchinson, M.F. and T. Jovanovic (1988). Niche analysis and tree species introduction. **For. Ecol. Manage.** 23: 47-59.

Busby, J.R. (1991). BIOCLIM - A Bioclimate Analysis and Prediction System. pp. 64-68. in **Nature Conservation: cost effective biological surveys and data analysis**. (eds. C.R. Margules and M.P. Austin). CSIRO, Melbourne. 207 pp.

CRES (1991). **Annual Report - Centre for Resource and Environmental Studies**, The Australian National University, Canberra.

Davey, S.M. and T.W. Norton (1990). State forests in Australia and their role in wildlife conservation. **Proc. Ecol. Soc. Aust.** 16: 323-345.

Hutchinson, M.F. (1987). Methods for generation of weather sequences. pp. 149-157 in **Agricultural Environments: characterization, classification and mapping**. (ed A.H. Bunting). CAB International, Wallingford, UK.

Hutchinson, M.F. (1989). Calculation of hydrologically sound digital elevation models. pp. 117-133 in **Proceedings Third International Symposium for Spatial Data Handling**. International Geographical Union, Columbo, OH.

Hutchinson, M.F, and T. Dowling (1991). A continental hydrological assessment of a new grid-based digital elevation model of Australia. **Hydrol. Proc.** 5: 45-58.

Lindenmayer, D.B., Nix, H.A., McMahon, J.P., Hutchinson, M.F. and M.T. Tanton (1991). The conservation of Leadbeater's Possum, *Gymnobelideus leadbeateri* (McCoy): a case study of the use of bioclimatic modelling. **J. Biogeog**. 18: 371-383.

Mackey, B.G. (1991). **The spatial extension of ecological relations in vegetation studies: a case study in tropical forests of north-east Queensland.** Unpublished Ph.D Thesis, The Australian National University, Canberra.

Mackey, B.G. and A. Bayes (1989). A modelling framework for the spatial extension of ecological relations in vegetation studies. pp. 93-97 in **Proceedings of the Eighth Biennial Conference of the Simulation Society of Australia**. IMACS, Canberra.

Mackey, B.G., Nix, H.A., Hutchinson, M.F. McMahon, J.P. and P.M. Fleming (1988). Assessing representativeness of places for conservation reservation and heritage listing. **Environ. Manage**. 12: 501-514.

Mackey, B.G., Nix, H.A. Stein, J.A., Cork, S.E. and F.T. Bullen (1989). Assessing the representativeness of the Wet Tropics of Queensland World Heritage Property. **Biol. Conserv**. 50: 270-303.

Margules, C.R. and J.L. Stein (1989). Patterns in the distribution of species and the selection of nature reserves: an example from *Eucalyptus* forests in south-eastern New South Wales. **Biol. Conserv**. 50: 219-238.

Moore, I.D., Grayson, R.B. and A.R. Ladson (1991). Digital terrain modelling: a review of hydrological, geomorphological and biological applications. **Hydrol. Proc**. 5: 3-30.

Neave, H.M., Cunningham, R.B., Norton, T.W. and H.A. Nix (1991). Evaluation of field sampling strategies for estimating species richness by Monte Carlo methods. pp. 466-471 in **Ninth Biennial Proceedings of the Simulation Society of Australia**. IMACS, Gold Coast.

Nix, H.A. (1986). A biogeographic analysis of Australian elapid snakes. pp. 4-15 in **Atlas of Elapid Snakes of Australia** (ed. R. longmore), Australian Flora and Fauna Series No.7 AGPS, Canberra, Australia.

Nix, H.A. and A.N. Gillison (1985). Towards an operational framework for habitat and wildlife management. pp. 39-55 in **Wildlife Management in the Forests and Forestry-Controlled Lands in the Tropics and Southern Hemisphere** (ed. J. Kikkawa). University of Queensland Printery , Brisbane, Australia.

Nix, H.A. and M. Switzer (1991). **Atlas of Tropical Rainforest Vertebrates of Australia**. Kowari Series No. 1. Australian National Parks and Wildlife Service and AGPS, Canberra.

Norton, T.W. and I.D. Moore (1991). Ecological applications of GIS and hydrological modelling: identification and management of Koala habitat in south east Australia. pp. 637-648 in **Proceedings of International Hydrological and Engineering Conference**, Perth, November.

Norton, T.W. and H.A. Nix (1991). Application of biological modelling and GIS to identify regional wildlife corridors. pp. 19-26 **in Nature Conservation 2: The Role of Corridors**. (eds. D.A. Saunders and R.J. Hobbs). Surrey Beatty and Sons in association with CSIRO, Sydney.

Norton, T.W. and J.E. Williams (1990). Wildlife management in Australia: new developments and opportunities using computer-based generic models. **Math. Comp. Simul**. 32: 231-235.

Norton, T.W., Mackey, B.G. and D.B. Lindenmayer (1990). Comments on biological and environmental data sets required for the Australian National Forest Inventory. **Aust. For**. 53: 88-95.

Norton, T.W., Neave, H.M., Nix, H.A. and I.D. Moore (1992). Approaches to forest survey, inventory, evaluation and planning using biophysical models of landscape processes. **Conference Proceedings, GIS '92**, Vancouver, British Columbia, Canada.

Richards, B.N., Bridges, R.G., Curtin, R.A., Nix, H.A., Shepherd, K.R. and J. Turner (1990). **Biological Conservation of the South-East Forests**. AGPS, Canberra, Australia.

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Forest monitoring using temporal GIS techniques

**Gail Langran**

Intergraph Corporation
2051 Mercator Drive
Reston, VA, USA 22091-3413

## Abstract

*Temporal components of forest monitoring include agents of change, and changes in forest state. To monitor forest resources and to determine the relative success of management strategies, both temporal components should be tracked and linked within a geographic information system (GIS). A desire to track silvicultural activities has focused attention recently on temporality in forest GIS. But it is not enough to track activities alone. The context in which activities are planned and performed—i.e., the state of the forest at those times—may justify actions that later seem questionable due to situational changes. Likewise, changes in forest stands are a useful basis on which to judge the relative success of activities performed. Stands that are managed one way can be compared to stands managed differently, after adjusting for unplanned or uncontrollable variations. The growing investigative area of temporal GIS has produced a number of techniques that could lead to a truly temporal forest GIS. This paper will discuss those techniques and how they are adaptable to forest management concerns.*

## Introduction

If a map is a model of reality, then a GIS is a greatly enhanced and empowered model of reality. However, a GIS that cannot model a chain of events and the changes it effects still falls far short of being a trustworthy basis for important environmental and economic decisions.

Temporal GIS (TGIS) is, as yet, a set of concepts and techniques rather than a commercial off-the-shelf system (for an overview, see Langran 1989a, 1990b). Would-be designers of a one-size-fits-all TGIS are challenged by the diversity of prospective users' needs. Until these diverse requirements are reconciled, it is productive to examine each application in depth to discover what specific uses it may have for a TGIS and how its needs might be met most effectively.

Forest monitoring obviously would gain from the availability of spatiotemporal data storage and analysis methods. The primary purpose of this discussion is to highlight some of the distinctive TGIS needs of foresters and some possible means of meeting them. A secondary purpose is to examine futuristic needs for spatiotemporal data analysis in forestry so the focus is not so limited that obsolescence outpaces cost-effectiveness. This is a serious risk. In the 1980s, organizations often defined their data capture needs so narrowly that the same data needed to be recaptured with better resolution, accuracy, or attribution within a matter of years. Since the essence of a TGIS is that temporal data are captured continually during routine updates (see USFS 1991 and Howard and Barr 1991), the nature of TGIS procedures, particularly those involving updates, ultimately shape the spatiotemporal data resource available within an organization. With careful planning, routine updates produce a spatio-

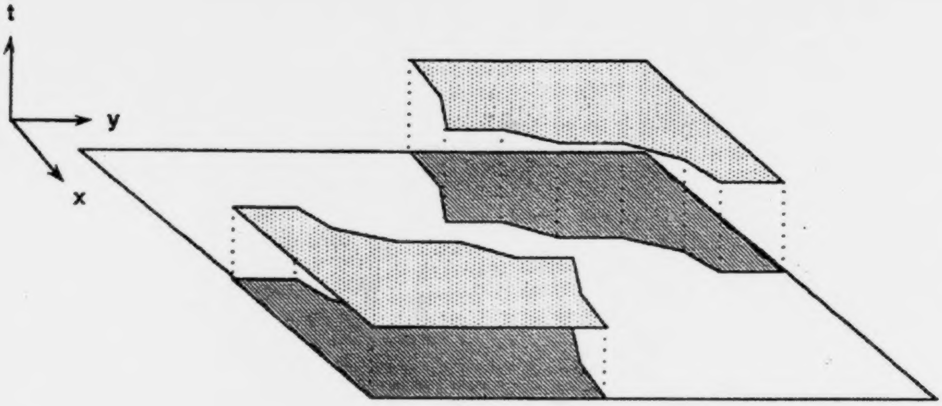


**Figure 1:** *A TGIS stores change data as amendments to a base state. Shaded areas in the SW and NE corners of the pictured region have undergone change, while the white area is unchanged.*

temporal data resource spanning generations of forest stands and foresters.

The discussion that follows briefly compares a temporal to an atemporal GIS, then considers the distinctive elements and emphasis of a forestry TGIS that can track activities and meet the spatiotemporal needs of the future. Some approaches to activities tracking that now exist within the forestry community are examined. The final section concludes with a list of goals for a forestry TGIS to challenge the system designer.

## Temporal vs. atemporal GIS

The essence of a TGIS is that it can store spatiotemporal data and analyze changes to spatial information over time. This means that updates are performed regularly and outdated information is treated as a valuable resource rather than a waste product.

Updates should correspond to when real changes occur in the world or when new information becomes available rather than when a new product is needed or when time permits. Only changed data are altered while unchanged data are left untouched. Conceptually, then, a TGIS describes changes as in Figure 1.

The capabilities of a TGIS enhance those already available in an atemporal GIS (Table 1). Since no off-the-shelf TGIS yet exists, analytical commands also are not standardized. However, a TGIS would help a user attain the answers to such questions as are listed in Table 2.

## Characteristics of forest GIS temporality

The most distinctive requirement of a forestry TGIS is that two sorts of temporality must be tracked: events that may cause change, and changes to the forest state itself. Events include silvicultural activities such as fertilization, spraying, harvesting, and replanting; wildfires; insect infestation; and unusual weather, such as windstorms, heavy rain or snow, or drought.

Events have spatial and temporal location and extent. Temporal location is within or spanning the past, present, or future, depending on whether the event is complete, underway, or planned. Temporal extent spans the period when an event is performed. Spatial location and extent are defined by the area in which the event occurs. Thus, to describe an event requires two dates (for start and finish), the coordinates needed to specify ground location, and descriptive

**Table 1:** *The capabilities of a temporal GIS include analysis, updates, quality control, scheduling, animation, and static mapping.*

| | |
|---|---|
| **Inventory** | Store a complete description of the study area, and account for changes in both the living world and computer storage. |
| **Analysis** | Explain, exploit, or forecast a region's components and processes. |
| **Updates** | Supersede outdated information with current information |
| **Quality Control** | Evaluate whether new data are logically consistent with previous versions and states |
| **Scheduling** | Identify or anticipate threshold database states, which trigger action by the system. |
| **Animation** | Display a dynamic summary of regional processes. |
| **Static Mapping** | Display spatiotemporal facts by traditional cartographic means. |

**Table 2:** Questions that could be answered by a TGIS. Temporal systems have tenses the present tense. In this case, the system could well require past, conditional, and future tenses.

| | |
|---|---|
| **Past** | How has a feature changed over its lifespan or during a given timespan? |
| | What was the state of an area as of a given moment? |
| | What was the state of an area when certain criteria existed? |
| | At a given moment, which features met certain criteria? |
| **Conditional** | *Given certain conditions...* |
| | ...how might a feature change during a given timespan? |
| | ...what might be the state of an area at a given moment? |
| | ...what might be the state of an area when certain criteria become true? |
| | ...which features may meet certain criteria at a given moment? |
| **Future** | What planned changes will occur to a feature during a future timespan? |
| | What is the planned state of an area at a given future moment? |
| | What is the planned state of an area when certain criteria become true? |
| | At a given future moment, which features are planned to meet certain criteria? |

information concerning type of event and details of the event. This familiar configuration of data suggests that for storage purposes, at least, events can be treated as any other geographic information.

Changes to the forest state can affect cutting units (or forest stands) or the forest infrastructure (e.g., roads, control points, etc.). Such changes may or may not form a causal relationship with events, so independent treatment is indicated. State changes also have temporal extent, which is delineated by surveys or other source documents that provide new information on the state. Temporal location is the timespan when a state held true. Spatial location and extent are the area on the ground where the state held true. Thus, to describe a forest state requires two dates (to define the effective dates of the state), ground coordinates, and descriptive information. Note that the data requirement is identical to events, although the data may be structured differently because of innate differences between events and states. Likewise, these are minimum data requirements. It is also useful to include several sets of dates to describe when the information became known and when the information entered the database; references to the survey or source that describes a state or event; and cross-references and comments.

Since forest state data describe a chain of events, each state borders the previous and next version, forming a set of temporal neighbors. For this reason, a given entity (for example, a stand) only needs to specify the start date of each version, since its finish date is automatically that of the next version. Also, if each state is associated with a survey that describes the discovery of that state, and each survey is associated with a date, the state inherits the date of the survey. Conceptually (and crudely), the major data components of a forest TGIS are events, forest stands, forest infrastructure, and surveys (Table 3).

**Table 3:** *The major data components of a forest TGIS.*

| **Component** | **Descriptors** |
|---|---|
| Event | Location, type of event, start date, finish date, source of info, details |
| Forest stand | Location, id of latest survey, key to previous version, details |
| Infrastructure | Location, id of latest survey, key to previous version, details |
| Survey | Location, survey id, details |

## Existing approaches for forestry TGIS

A number of ideas exist for maintaining some level of temporal data in a forestry GIS.Some have been implemented with success, others are untested approaches. All these efforts represent ambitious advances over an atemporal GIS. Therefore, the discussion is not intended to discredit these efforts but rather to examine their achievements and limitations, since new approaches often germinate in existing technology.

### Narrative description

Among the simpler solutions for describing activities performed on forest stands is to store the limits and descriptors of forest stands, then include a narrative description of activities occurring within the stand. For example, if brush were cleared in part of the stand, the database records for that stand would include a free-form comment stating "brush cleared from SW 5 acres in April 1991."

This method is a natural way to describe activities, and it relieves foresters of mapping limits of activities within stands and performing graphic edits to enter them in the system. Any forester examining the records for this stand would be able to retrieve the full set of narrative descriptions of activities performed. However, it would not be possible to sort them sequentially, to map them, or to perform computer analysis on this information. It does not build a data resource for future temporal analysis, nor does it exploit the computational powers of a GIS. In addition, it tracks only activities and does not maintain a history of forest stand temporality.

### Rasterized temporal data

Since the usefulness of temporal data is assumed but unproven, it is reasonable to economize when capturing and archiving temporal data. A Weyerhauser system described in (USFS 1990) takes a yearly snapshot of the forest state by rasterizing the vector data. The raster data can be backdropped against the vector data for visual comparison. And since raster GIS procedures are less computationally intensive than are vector, temporal analysis is also a seemingly simple matter of comparing pixels in different time slices.

If the value of temporal data is considered to be mainly archival, this system is an appropriate approach. Visual comparison of current and past data states also is well supported by this system. But while the previous (narrative) approach described activities without documenting changes in the stands, the rasterization approach captures ONLY changes within the stands and does not document activities performed. If that limitation could be overcome, this system still does not provide a healthy basis for a temporal GIS because:

1. Temporal data are captured yearly rather than when change becomes known, and changes are not referenced to the source of knowledge about them (i.e., a survey).
2. To perform temporal analysis by comparing several rasterized snapshots, precise registration of the snapshots would be critical. Without precise registration, there is no guarantee that stand histories are correctly assembled.
3. The snapshot approach uses storage inefficiently, since unchanged data are replicated redundantly.

### Space-time composite

Generating a space-time composite of spatial and temporal information is an method for tracking the temporality of forest stands that is used with minor adaptations by the U.S. Washington State Department of Natural Resources (Miller 1991, Miller 1988) and by the Canadian Forest Service in New Brunswick, B.C. (USFS 1990). Langran and Chrisman (1988) and Langran (1992) describe the conceptual basis for this approach, which is tailored to polygon data that exhaustively cover a region.

A space-time composite is a single polygon coverage generated from a snapshot series by overlaying the snapshots to define the greatest common spatiotemporal units, i.e., the polygons that share like histories of attribute change (Figure 2). An important advantage of the space-time composite is that it uses familiar concepts and can be implemented within existing commercial GISs, although development of special software modules atop the commercial software can improve performance (see Langran 1989b). Its major disadvantage is that over time, serious performance degradation can occur from two causes:

1. Changes create new chains and nodes by splitting existing objects, leading to successive fragmentation of the data.
2. When new chains and nodes are created through splitting, they must be referenced by existing objects which formerly referenced the parent chain and nodes. As more chains and nodes are generated through splitting, this need to renumber becomes increasingly burdensome.

Since a temporal GIS should be considered an enduring resource whose database grows and is usable indefinitely, performance degradation over time is a serious concern. If older data are archived, separate procedures for storage and retrieval must be developed because archives cannot reasonably be renumbered. While none of these difficulties makes this approach unworkable, all must be addressed somehow. One possibility is to delimit the temporal depth of a forest stand to one "generation," i.e., from harvest to harvest (see Vrana 1990 for a diagram of the forest harvest cycle, as described by Miller 1988). In other words, once a stand is harvested, the fragments that outline activities within the stand are dissolved or reconstituted into only those chains and nodes needed to describe the stand spatially. Unless special measures were taken, this would mean that the effects of activities beyond one generation could be analyzed.

### Feature histories

The essence of the space-time composite is to account for all space at all times, treating features as attributes of space. A more direct approach to describing feature data is to treat the features themselves as objects in the database and to record their histories as linked lists of database records. A method of recording the histories of features was developed by Intergraph when designing a spatiotemporal feature database for the U.S. National Ocean Service (see Langran 1990a for a description of this system's temporal treatment). When complete, this system will be used to trace changes to such hydrographic data as shoreline, buoys, and lighthouses.

To describe feature histories, the attributes that describe different versions of a feature are treated as separate database records, which are linked within a relational database by their effective dates (Figure 3). Geometric change is treated by referencing the correct spatial objects to each version of the feature as it changes over time (Figure 4). Thus, spatial objects accumulate

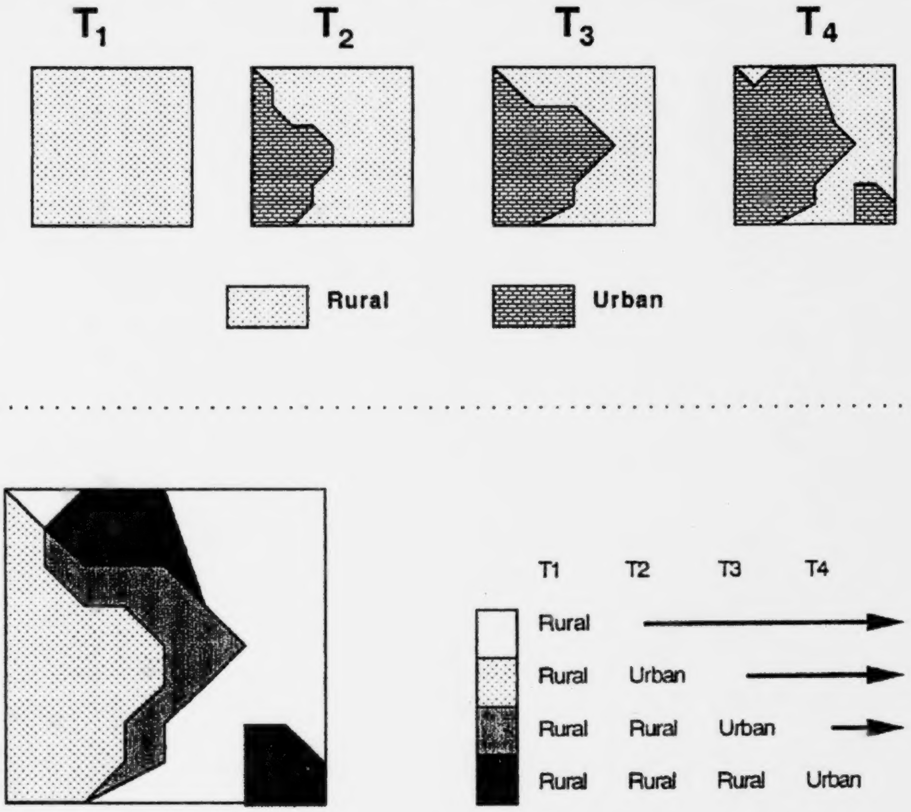


**Figure 2:** *Construction of a space-time composite. The snapshot sequence (shown at top) is overlaid to create the set of greatest common spatiotemporal units, i.e., polygons that share a common history.*

over time to describe all versions of all features, but the spatial objects of different time slices are not intersected with one another.

A major technical advantage of tracking feature histories is that past data structures are not affected by subsequent changes. This results in the following benefits over the space-time composite:

1. Objects need not be renumbered retroactively when present change affects past data structures.
2. Archives can be placed on read-only medium, since they are immune from subsequent change.
3. Data are intersected only with other data that are valid at that moment in time, which reduces the fragmentation.

This approach has been flexible enough to describe changes to point, line, and area features. While forestry data are somewhat distinctive (as described in the previous section), the method is likely to be extensible to this new purpose.

A second concept used for the National Ocean Service's system that will prove important to a fully temporal forest GIS is that of the source or survey catalog. This digital catalog stores the spatial boundaries of all sources of information and any attributes of interest, such as date of survey, survey method, horizontal and vertical datum, surveyor, method of data capture (in the field and in the computer lab, if both are of interest), and so on. All information gleaned from the survey references that entry in the catalog, so its validity can be evaluated at any time. The existence of a survey catalog means that this information can be mapped and queried independently of states and events.

The concept of the survey catalog is extensible to event and activity data. Then it would be possible to display and query the limits of activities within an area, and view the details of a particular activity by pointing to its boundary on the workstation screen. Likewise, it would be possible to overlay activity limits on stand limits to perform analysis, and to overlay event limits on survey limits to see if new surveys are needed to guage the effects of events. The possibility of treating activities as discrete data elements has been discussed by the Events Tracking Working Group of USFS Region 6 (USFS 1990).

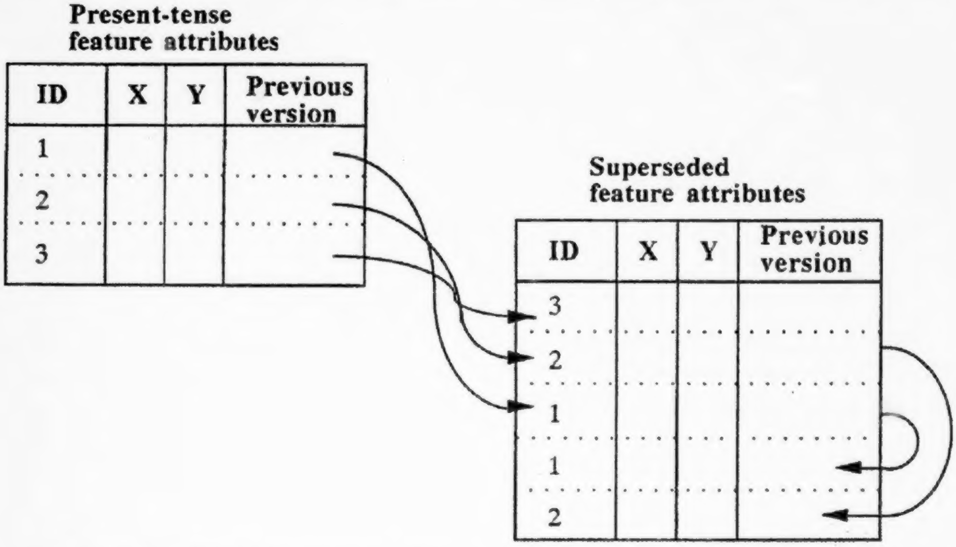


**Figure 3:** *Tracing attributes changes to features. Features 1, 2, and 3 all have past versions, which have been superseded and linked to the present-tense versions. Features 1 and 2 have two psat versions each, which are stored and linked in the superseded feature table.*

## Conclusion

Meeting immediate and urgent temporal data processing needs is usually possible by manipulating existing data processing procedures. However, creating an enduring and flexible temporal GIS for a given application requires care and thought to avoid premature obsolescence.

The forest TGIS options explored here all have merit, and the best possible approach may have ingredients from several. Among the primary goals to be met are:

- Track changes to forest stands AND events that may cause change.
- Describe planned, active, and completed events.
- Describe homogenous vegetative stands.
- Do not impact the efficiency of present-tense data analysis when past- and future-tense data exist.
- Develop routine but rigorous update procedures that cause the temporal data resource to grow naturally over time.
- Ensure that older data can be moved to less-expensive storage media (e.g., archived), while still being accessible for longitudinal analysis.
- Ensure that the system does not become less viable as the temporal data resource grows.
- Develop easy methods of answering spatiotemporal questions by developing standard temporal map formats and simple query-by-example forms.

## Acknowledgments

I would like to thank Dale Guenther of USFS Region 6 in Portland, whose personal interest in this topic led to several interesting conversations. These conversations gave me insights on forest TGIS requirements that went beyond what I could have gleaned from papers on the subject. My overview of suggested technical approaches to activities-tracking requirements is based on a set of notes from an events-tracking working group (cited in references).

## References

Freedman, David (1991). "The Forest for the Trees." CIO, January, 25-32.

Howard, David C. and Barr, Colin J (1991). "Sampling the Countryside of Great Britain: GIS for the Detection and Prediction of Rural Change." *Proceedings of GIS* '91, 171-176.

Langran, Gail and Chrisman, Nicholas R. (1988). "A Framework for Temporal Geographic Information." *Cartographica* 25, 3, 1-14.

Langran, Gail (1989a). "A Review of Temporal Database Research and its Use in GIS Applications." *International Journal of GIS* 3, 3, 215-232.

Langran, Gail (1989b). "Accessing Spatiotemporal Data in a Temporal GIS." *Proceedings of AUTO-CARTO* 9, Baltimore, MD, 191-198.

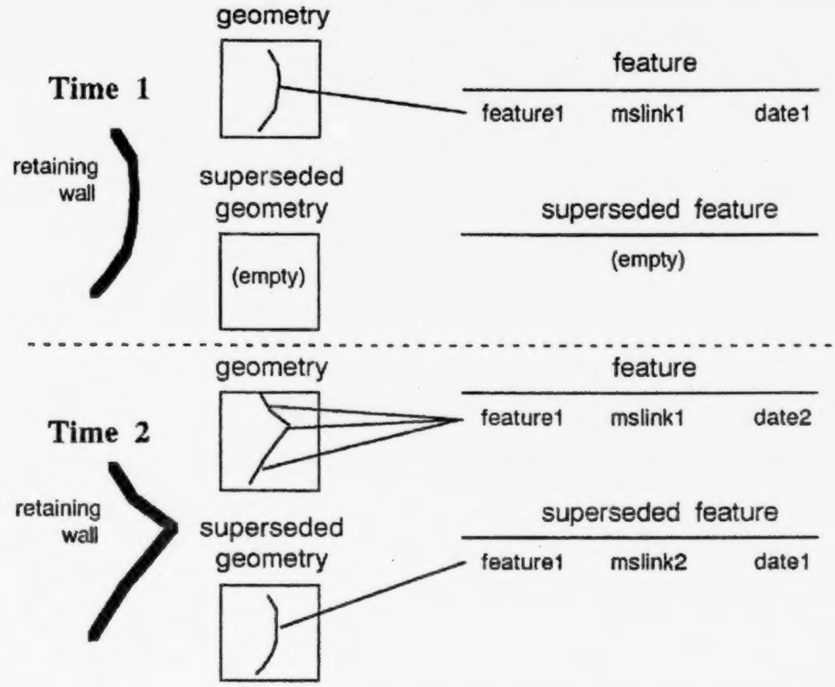

geometry
feature
Time 1
feature1 mslink1 date1
retaining wall
superseded geometry
superseded feature
(empty)
(empty)
geometry
feature
Time 2
feature1 mslink1 date2
retaining wall
superseded geometry
superseded feature
feature1 mslink2 date1

**Figure 4:** *Tracking geometric changes to features. When the retaining wall is rebuilt, its previous descriptors and geometric objects are superseded and replaced in the present tense with the newly effective objects.*

Langran, Gail (1990a). "Tracing Temporal Information in an Automated Nautical Charting System." *Cartography and GIS* 17, 4, 291-299.

Langran, Gail (1990b). "Temporal GIS Design Tradeoffs." *Journal of the Urban and Regional Information Systems Association* 2, 2, 16-25.

Langran, Gail (1991). "Dilemmas of Implementing a Temporal GIS." *Proceedings of the International Cartographic Association*, Bournemouth.

Langran, Gail (1992). *Time in Geographic Information Systems*. London: Taylor and Francis.

Miller, Steve (1988). Presentation on forest temporality given to a University of Washington graduate seminar on temporal GIS.

Miller, Steve (1991). "Washington DNR's Forest Management Tracking System: A Design for the Many-To-Many Relationship." *Proceedings of GIS* '91, 97-100.

U.S. Forest Service (1990). Unpublished notes from the Region 6 Activities Tracking Working Group meeting. Unpublished notes from a meeting held 18-19 July in Clackamas, Oregon.

U.S. Forest Service (1991). "Geographic Information System Database Management Plan." Region 8 GIS Task Force. Atlanta: USDA Forest Service Southern Region, 17 June.

Vrana, Ric (1989). "Historical Data as an Explicit Component of Land Information Systems." *International Journal of GIS* 3, 1, 33-49.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Analysis of the spatial distribution of a forest population for the planning of a forest inventory

**Jean-Marie Bilodeau, QFE., M. Sc.**

Chaire industrielle en géomatique appliquée à la foresterie
Yvan Bédard, QLS Ph. D., Director,

**Kim Lowell, Ph. D.,**

Co-director, Chaire industrielle en géomatique appliquée à la foresterie
Centre de recherche en géomatique
Faculté de foresterie et de géomatique
Université Laval
Local 0611, pavillon Casault
Ste-Foy (Québec)
Canada, G1K 7P4
Téléphone: (418) 656-7894
Fax: (418) 656-7411

## Abstract

*GIS technologies have provided new possibilities concerning knowledge and management of forest land. One activity that will be influenced by GIS tools is the planning of forest inventories. This activity tends to obtain information about a population that has not only a distribution relative to descriptive attributes (species, density, height, etc.), but also spatial characteristics and a spatial distribution (such as grouped or regular). The analysis of the spatial distribution of a population will show the spatial pattern of forest stands. Respecting the spatial characteristics and the distribution of a population in the planning of a forest inventory could be useful for positioning samples to provide more representative forest information.*

## Résumé

*L'arrivée des SIG dans le domaine forestier ouvre la porte à de nouvelles avenues en matière de connaissance et de gestion de territoire. Un des processus qui est susceptible de subir des transformations importantes est celui de l'inventaire forestier. Ce processus vise à obtenir de l'information sur une population qui se distribue selon certains critères descriptifs (essence, densité, etc.) mais qui comporte également des caractéristiques spatiales ainsi qu'une distribution spatiale (groupé, régulier, etc.). L'analyse des caractéristiques spatiales et de la distribution spatiale d'une population apporte des précisions tant qu'à l'agencement particulier selon lesquels les peuplements forestiers se distribuent sur le territoire. La thus will put more samples in the other strata (which are already sampled adequately). This situation causes a change in the number of samples per stratum, relatively to the number calculated during inventory planning. In order to be able to respect the number of samples per stratum, we must have a good spatial distribution to respect local variations of the population.*

*We think that this situation can be rectified if we respect the spatial characteristics and the distribution of the population, and if we use a GIS to plan the forest inventory.*

## Objectives

The primary objective of our research is to determine a way to analyse the spatial characteristics and distribution of a population. The secondary objective is to improve the planning of forest inventories by using the knowedge of the spatial characteristics and distribution in conjunction with a GIS.

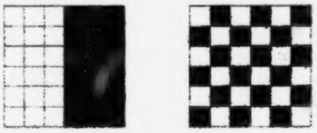
## Methodology

To achieve these objectives, we analyzed the problems inherent in the planning of a forest survey. We identified the causes of these problems and developed a theory that can help us to describe spatial distribution of a population. This theory has been applied to the planning of a forest inventory.

## Spatial chararteristics and distribution of a population

The first step in the study of the spatial characteristics and distribution was to examine the scientific literature. In our study, we defined the spatial distribution of a population (or sub-population) by the spatial characteristics between the objects which define its spatial pattern and location. These characteristics can be observed in one, two or three dimensions. In our application we consider only the two dimensions.

To begin our study of spatial characteristics, we used the concept of spatial autocorrelation (SA) which is clearly described by Odland (1988). Figure 1 presents two kinds of spatial autocorrelation. First, the objects have a tendency to be grouped, in a positive SA, second, the SA coefficient is negative as objects have a tendency to be dispersed.



**Positive** **Negative**

**Figure 1**

Types of spatial autocorrelation

Applying these ideas to a forest population, we can determine if forest stands of a particular stratum are spatially regrouped, systematically distributed, or randomly distributed. With one measure of SA, we can also evaluate whether or not two or more strata have a tendency to be adjacent. Many techniques can be used to measure spatial autocorrelation. The most frequently-used methods are the join-count-statistic, Moran's I and Geary's R (Odland, 1988). The method chosen depends on the scale of the measurement (nominal, ordinal, interval, ratio).

After this analysis, we examined the spatial reference of one object (Chevallier et Bédard 1990), and attempted to apply these notions to an entire population. If we draw a line around all elements of a population, we can evaluate the population as a single object. In this way, we can evaluate the position, dimension, orientation and shape of the population.

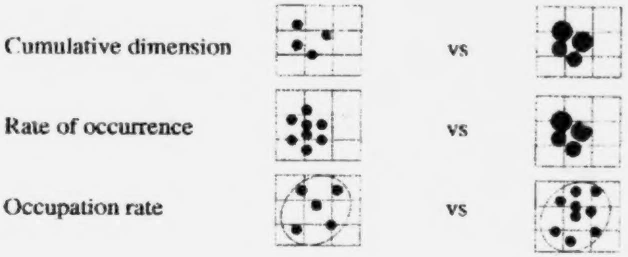


**Figure 2**

Spatial characteristics of a population

The dimension of a 2-D object, such as a population is the area. In the case of a population, composed of several objets, it can be described in greater details. The dimension of a population can be considered to be the sum of the areas of all objects that compose the population. In Figure 2, the dimension of the population is the total area of the ellipse around the population. To make the distinction we call the first one the cumulative dimension (sum of the area of each object), and the second one, the dimension of the population (total area of the ellipse around all objects). The latter description is linked to notions of areal distribution.

Depending on the nature of the population, we can conduct additional analysis on the number of objects per cumulative dimension or per dimension of the population. To visualize these notions, see Figure 3.

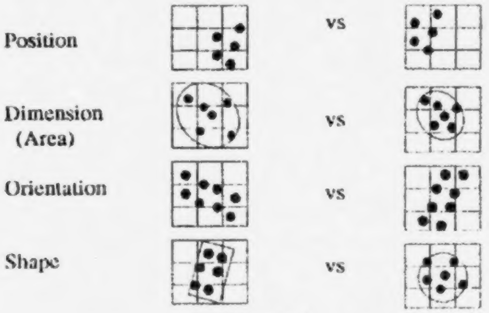


**Figure 3**

Representation of spatial characteristics of a population

In this kind of analysis, we can utilize the notion of spatial density which can be defined as the occupation rate divided by the rate of occurrence. This equation can be simplified and to produce the definition of a spatial density as the cumulative dimension divided by the dimension of the population.

All these characteristics may or may not be easily defined depending on the type of the spatial distribution which we have in our population. For example, shape is easy to determine if the spatial distribution is regrouped. Conversely, if we have a random distribution, it is difficult to determine the shape and/or the orientation. In addition, it may sometimes happen to have mixed spatial distribution.

We can also obtain some information about the spatial distribution if we overlay the population with a grid or some other spatial subdivision. Then, we can analyze rate of occurrence and occupation rate per subdivision or the dimension of the population and cumulative dimension.

## Application to forest inventory planning

To see how these ideas can help us in the planning of a forest inventory, we chose a territory (1200 ha) that is part of the Montmorency forest, the research forest of Laval University, that we have on a GIS. After the realization of the first steps of the planning (identification of objectives, territory, variables and variance), we were able to describe the spatial distribution of the forest population. We began with a measure of spatial autocorrelation (SA) using the regrouped strata. With the join-count-statistic (categorical data) we checked if the overall spatial distribution was grouped, regular or random. After this first test, we utilized the volume of each forest stand instead of its (categorical) type designation. With continuous data, it was then possible to use Geary's R and Moran's I to measure SA. All of these tests suggested that the spatial distribution of our forest population was random.

As mentioned, with a random distribution, it is difficult to estimate some spatial characteristics of the population (e.g., orientation, shape). However we have other spatial characteristics with which to describe the population.

If we recall the problems in forest inventory, in order to respect the spatial distribution of the population we should have at least knowledge of the position of the population. If we establish a grid on the territory, we can analyze the cumulative dimension of the population for each stratum and for each grid partition. If we want to respect local variation within a population, samples within each stratum must be proportional to the cumulative dimension for each stratum for each division. In this way, it is easier to respect the statistical distribution of samples by stratum that was defined initially. Thus we have defined a spatial subdivision that could also respect the operational constraints of a forest inventory.

Many spatial constraints exist. One of the most important constraints is the necessity to regroup 6 to 8 samples on a transect. This operation results in many restrictions on the positions of the samples. If we want to continue to meet this constraint we have to chose a subdivision that allows us to respect it. After some tests, it was determined that the optimal subdivision is the compartment. This subdivision is equivalent to the Parcel within the public forest in Quebec.

## Operational procedures

After doing this analysis of the descriptive and spatial distributions, we began to make the forest inventory plan. The first step was to determine the number of samples per stratum for the entire population. We did it by analyzing the variances from the previous plan. Table 1 is an example of this.

**Table 1. Number of samples by stratum according to a conditional distribution**

| Strata | Number of samples |
|---|---|
| SS A4 30 I | 4 |
| SS A4 30 II | 1 |
| SS A4 30-50 I | 7 |
| SS B4 30 I | 2 |
| SS B4 30 II | 2 |
| SS C4 30 I | 1 |
| SS B3 50 I | 4 |
| SS B3 50 II | 6 |
| SS C3 50 II | 3 |
| SS B3 70 I | 2 |
| SS B2 70 II | 1 |
| BBSR AB34 30-50 I-II | 3 |
| CT 79 | 1 |

*Note : SS represents species, A4 is a classe of density and height combination, 30 and 50 are age (years), and I-II are classes of site quality.*

With these subdivisions, we computed the percent area for each stratum for each subdivision. We took this percentage and multiplied it by the number of samples for each stratum. This produced a table that shows the number of samples for each stratum in each subdivision.

**Annex 1.** Forest inventory plan in the Montmorency forest

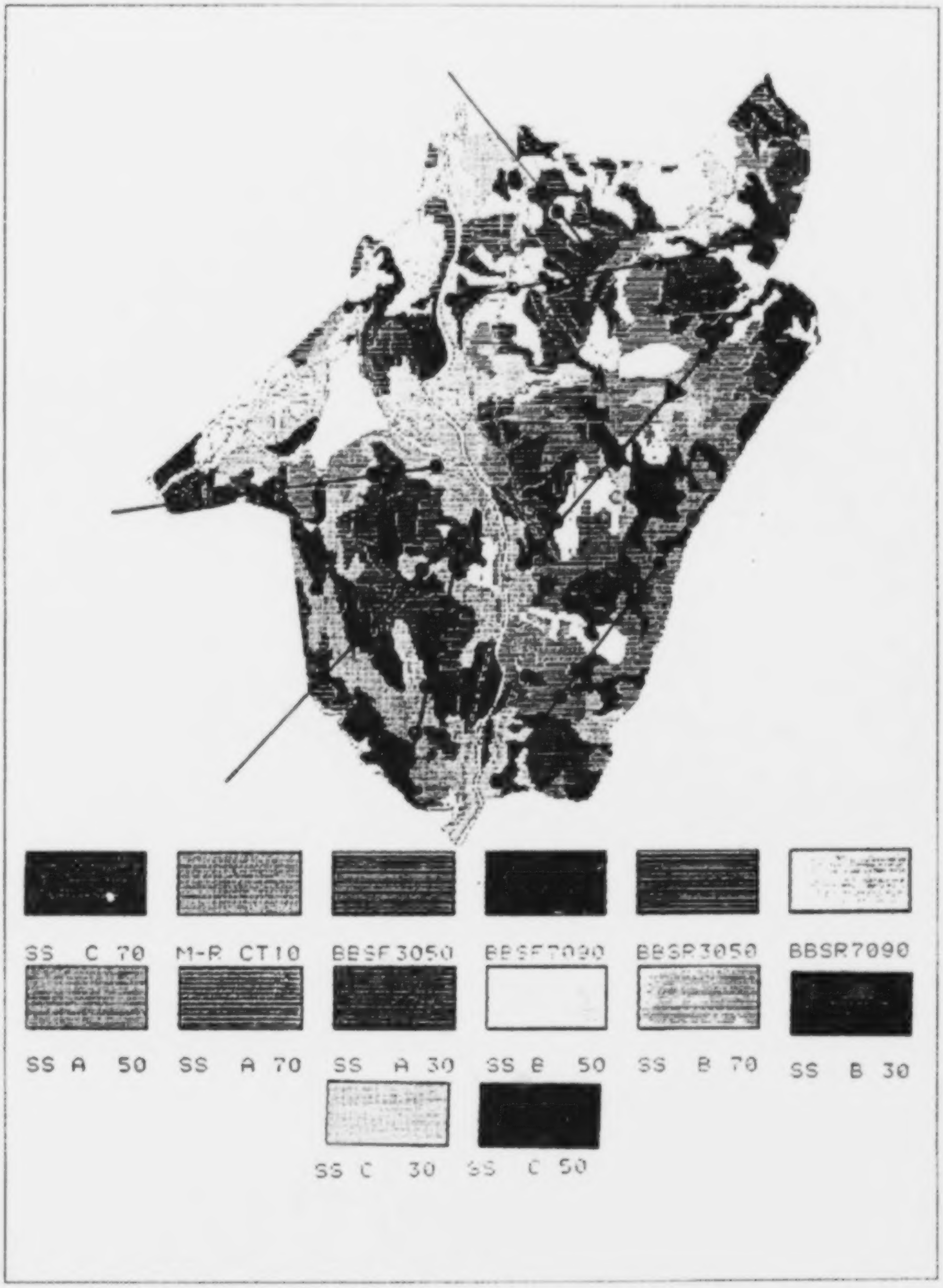

SS C 70
M-R CT10
BBSF3050
BBSF7090
BBSR3050
BBSR7090
SS A 50
SS A 70
SS A 30
SS B 50
SS B 70
SS B 30
SS C 30
SS C 50

**Table 2.** Distribution of samples by strata by subdivisions

| Compartment | Strata | Number of samples |
|---|---|---|
| 13 | SS B4 30 I | 1 |
| | SS B4 30 II | 1 |
| | SS C4 30 I | 1 |
| | SS B3 50 I | 1 |
| | SS B3 50 II | 2 |
| | SS C3 50 II | 1 |
| | BBSR AB34 30-50 I-II | 2 |
| 14 | SS A4 30-50 I | |
| 2 | | |
| | SS B3 50 I | 1 |
| | SS C3 50 II | 1 |
| 15 | SS B3 50 II | 1 |
| 23 | SS A4 30-50 I | |
| 5 | | |
| | SS B3 50 I | 1 |
| | SS B3 50 II | 2 |
| | SS B3 70 I | 1 |
| 24 | SS A4 30 I | 1 |
| | SS B4 30 II | 1 |
| | SS B3 50 I | 1 |
| | SS B3 70 I | 1 |
| 25 | SS A4 30 I | 3 |
| | SS A4 30 II | 1 |
| | SS B4 30 I | 1 |
| | SS B3 50 II | 1 |
| | SS C3 50 II | 1 |
| | SS B2 70 II | 1 |
| | BBSR AB34 30-50 I-II | 1 |
| | CT 79 | 1 |

The information in Table 2 can help us to respect the spatial characteristics and distribution of the population and to respect the number of samples per stratum for the territory. The second tool required is a GIS with a thematic map that shows different strata. With the results in table 2 and a thematic map, it is relatively easy to plan a more representative forest inventories, i.e. which is less sensible to local, non-measured variations (the annex 1 presents the result of the plan).

## Observations and conclusion

In the usual method, it is difficult to respect statistical distribution and local and unknown variation. With this new approach and a GIS, we can analyze spatially the population in the planning of a forest inventory, and can increase the efficiency of the forest survey.

This exercise showed that is easier to locate the sample in a forest stand for each strata per subdivision. We begin by locating the sample in the corresponding color and, after that, by trying to regroup samples on a transect. In our tests, we maintained the statistical distribution of the number of samples per stratum and also respected the spatial distribution of the population. This means that if the population or subpopulation has local variations, the results of our planning should account for this. We believe that this method could help to achieve more representative forest inventory results. Knowledge of the spatial distribution of the forest population can also help us to determine the location of each sample with respect to the spatial distribution of the population.

## Literature cited

Odland, J.1988, Spatial Autocorrelation. Scientific Geography Series, Vol 9, Department of Geography, Indiana University, Editor: Grantlan Thall, United States, 87p.

Chevallier, J.J. et Bedard, Y., 1990, *Classification des types de reference spatiale utilisee dans les systems d'information a reference spatiale (SIRS)*. Universite Laval, Centre de Geomatique, Departemant des Sciences Geodesiques et de Teledetiction, notes de cours.

G I S
92

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Quality control in data base development and spatial analysis

**Brendan G. Mackey**

**Daniel W. McKenney**

Forestry Canada, Ontario Region
Sault Ste Marie, Ontario P6A 5M7

## Abstract

*This paper discusses some issues relating to data base development and the spatial analysis of biological site data. The context for this discussion is an ecological economics study based in the forested lands of Ontario. Emphasis is given here to the problem of quality control at all levels of data integration. Procedures are suggested to assist in determining data error and variability.*

## Introduction

Our current work is taking place in the context of Ontario. Forestry Canada has provided money, through the Green Plan, for a project that essentially involves the examination of potential production opportunities and trade-offs between certain ecological values and wood production in Ontario forests. The ecological "values" of interest include representativeness, habitat suitability, diversity and naturalness.

The three major stages of the project are identified in Table 1. The first stage is the development of a primary data base such that ecological and wood production possibilities can be modelled spatially. This data base will include the climatic, terrain and parent material attribute data needed to model (or approximate) the major physical determinants of biological processes. These are represented by the thermal, radiation, moisture and mineral nutrient regimes. Extensive use will also be made of biological site data (the distribution and abundance of taxa, productivity data and community structure). The second stage will involve the statistical analysis and spatial extension of the site-based data using the primary attribute data as predictors. Data from thematic and remotely sensed sources will also be utilized at this stage. The third stage will involve examining the trade-offs between the different production possibilities. This will require analyzing the preferences of both decision makers and the public.

The project requires various levels of spatial analysis and data integration from a variety of sources. This is common to all GIS applications. The primary attribute data (e.g. climate) are analyzed to derive secondary attributes (e.g. moisture index) which in turn are used to model tertiary attributes such as future timber supply or biodiversity. Like most researchers we are keen to progress to the second and third stages of the project where many interesting problems remain to be resolved. Extensive and elaborate analyses however are to no avail unless rigorous quality control procedures are adopted at all stages. This suggests a need for information about the error and variability associated with spatial data and predictions at all levels of integration: primary; secondary and tertiary. The remainder of this paper discusses some issues relating to quality control through the various stages of the project, and examines procedures to assist in determining data error and variability.

**Table 1. Three stages of Ontario forest trade-off evaluation project**

| | |
|---|---|
| **Stage 1** | - development of primary data base |
| | : terrain, climate, parent material, landcover |
| | : biological site data, eg. forest plot data |
| **Stage 2** | - generation of spatial estimates for selected ecological and timber indices |
| | : eg. biodiversity, representativeness, growth and yield, habitat potential |
| **Stage 3** | - analysis of the covariance and conflict between the various ecological and timber indices through space and time |

**Table 2. Source data for data base development**

| | |
|---|---|
| (a) | **thematic maps** |
| | eg. geology, soils, vegetation types from air photo interpretation |
| (b) | **remotely sensed digital data** |
| | eg. Landsat TM |
| (c) | **digital climatic models** |
| | eg. SPLINB modelling procedure |
| (d) | **digital terrain models** |
| | eg. ANUDEM modelling procedure |
| (e) | **site data** |
| | eg. species location data, forest plot growth and yield data |

## Quality control in primary data base development

We suggest that the limiting factor for spatial analysis is no longer the availability of suitable analytical techniques or simulation algorithms. Rather, the value of GIS for biophysical modelling hinges upon the accuracy and precision with which spatial and temporal estimates of the primary environmental regimes can be generated. The utility of the spatial analyses and modelling undertaken in Stages 2 and 3 of our project will depend largely on the quality of the primary data base. Therefore it is critical that firm control is maintained on data quality and the variability associated with a given spatial estimate.

Different source data can be drawn upon in the primary data base development phase. These are listed in Table 2. Air-photo interpretation (API) is the technique traditionally used for deriving spatial estimates of field data. A common approach in data base development is to provide aerial coverage of a thematic layer by digitising the mapped polygon boundaries and accompanying descriptive attribute data. GIS systems such as ARCINFO have been developed largely on the assumption that existing thematic maps, derived from API, are the principal data source.

Thematic maps have a role in modelling the primary environmental regimes, though we would suggest a diminishing one. One reason for this is that many of the parameters required to estimate these regimes are simply not amenable to API. Dynamic processes, for example, involve estimating or monitoring the flow of resources or energy through space and time (eg. climatic averages, available moisture). Consequently a greater emphasis is required on the remaining data sources (remotely sensed spectral and thermal, digital climatic and terrain, site data). These data sources also lend themselves more readily to statistical verification procedures.

### (1) Spatial extension of climatic averages

An excellent example of the extent to which quality control can be applied in data base development is provided by the climate interpolation procedures of M.F. Hutchinson and colleagues (see Hutchinson and Bischoff 1984, Hutchinson 1987). The procedures (which centre on the computer program SPLINB) generate mathematical surfaces that describe the spatial distribution of mean monthly climate across a landscape. The surfaces are derived from interpolation of long-term weather records collated at a network of sites. The interpolation procedure utilizes the Laplacian smoothing spline methods of Wahba and Wendelberger (1980). A significant development has been the incorporation of independent variables in addition to latitude and longitude. This enables known physical dependencies of precipitation and temperature on elevation to be taken into account. In the case of radiation, the third independent variable is a cloudiness index.

The surface fitting procedure uses a generalized cross-validation technique to iteratively examine the error associated with fitting each data point. Thus for each site an estimate is generated of the climatic variable when that site is not used to derive the surface. This can be compared to the actual climatic averages at each site. The root mean square error associated with fitting the surface at each site is calculated and provides a useful diagnostic for identifying either data errors (eg. incorrect geocodes) or geographically isolated sites where there is inadequate data to provide a better fit. Additional statistics generated by the procedure also serve as useful diagnostics, and the overall error associated with the fit is determined. The diagnostics assist in finding the balance between fitting the surface more or less tightly to either 'good' or 'bad' data sites. The entire process is open ended in that the surfaces can be re-calculated should additional site data become available.

The procedure has been applied to a variety of landscapes including the Australian and African continents, China and New Zealand. We are applying them to Canadian climatic data as part of a collaborative study between Forestry Canada, The Centre for Resource and Environmental Studies at the Australian National University, Don MacIver and colleagues of the Atmospheric Environment Service of Environment Canada, and Denis Joyce and colleagues at the Ontario Ministry of Natural Resources. Preliminary analysis for Ontario has generated surfaces with an error of about 8% for long-term mean monthly precipitation and 3% for minimum temperature.

### (2) Digital terrain modelling

Digital elevation models are a fundamental component of a primary spatial data base. Digital topographic data have become available for much of North America and, at varying scales, are available for many parts of the world. The problem however is to efficiently process these data into a digital elevation model (DEM), ie. a digital representation that retains specific terrain characteristics. The desired characteristics will vary with the purpose of the study. For example, gridded estimates of climate can be obtained by coupling a DEM to the climate interpolation surfaces described above. When developing a DEM for this purpose it is important that all the high points in the landscape are sampled as elevation-controlled lapse rates are a major determinant of local climate. However if the aim is to model surface flows or water movement through a catchment then a DEM must be hydrologically sound. This means that there must be no spurious sinks or heights (local maxima and minima), and that the drainage network is adequately reproduced.

The ANUSPLIN computer program (Hutchinson 1989) provides the basis of a procedure for developing hydrologically sound DEMs from digital topographic data. A drainage enforcement algorithm is used to automatically remove spurious sinks, and diagnostics are generated that enable the user to determine how well the drainage characteristics of the terrain have been captured. The procedure accepts contour strings, spot heights, sinks and stream lines as input data, and outputs a regular grid of elevations at a user-defined grid resolution and format.

## Quality control for secondary and tertiary spatial data

### (1) Secondary spatial data

The problem of quality control persists in the second and third phases of our Ontario project where the primary data will be used to derive secondary and tertiary spatial predictions. For example, the digital terrain and climate models theoretically provide many of the basic data needed to estimate the availability of moisture at any point in a landscape. However a definitive calculation needs to integrate the effects of precipitation, radiation, soil moisture storage, upslope contributing area, as well as the effects of slope, aspect, horizon and land cover on the local evaporative regime.

Defining the local evaporative regime is critical and quickly becomes problematic when dealing with large areas. The estimates of radiation generated by the climate interpolation procedure described above are based on data located on horizontal planes. Mackey and others (1988) describe a technique for obtaining gridded estimates of radiation using a computer program CLOUDY to incorporate the effects of slope and aspect (see also Austin and others 1984). A problem remains however in applying these to gain improved spatial estimates of evaporation (eg. by using the modified radiation term as input to a modified Penman evaporation equation). The terrain/radiation modification procedure based on CLOUDY uses radiation averaged over a daily time integral. The effect of inclined surfaces however results in variation to the diurnal distribution of both radiation and temperature. The timing of these changes is not necessarily coincident. Evaporation will only be increased on an inclined surface when both temperature and radiation are increased and their diurnal peaks coincide. Therefore in order to model the effects of inclined surfaces on evaporation it would be necessary to simulate the diurnal temperature and radiation curves, and the spatial distribution in the timing of their maxima.

An example of a simple water balance equation is WATBAL (Fitzpatrick and Nix 1970). Input attribute data include monthly estimates of precipitation, evaporation and soil available moisture. Once again, a moisture index derived using this equation is limited by the spatial quality of these various input parameters. The most difficult parameter to model is perhaps available moisture as this depends on obtaining spatially reliable estimates of, as a minimum, soil texture and depth of profile (see Murtha 1988). For small sub-watersheds this is feasible, but the calculation becomes difficult when dealing with extensive areas such as the forested lands of Ontario where such data are generally not available.

Digital elevation models can be used to generate indices of various catchment processes simply based on accumulated upslope catchment area, slope and aspect (eg. Moore and others 1991). Soil moisture storage and evaporation can be factored into these equations but once again the problem remains of obtaining spatially reliable estimates across extensive areas for the additional parameters.

### (2) Tertiary spatial data

Both primary and secondary attribute data can be used to analyze the environmental relations of biological phenomena. The advent of computer-based methods for modelling geographic data provides the foundation upon which new solutions to the problem of spatially extending these site data are being built.

Algorithms can draw upon either empirical, deterministic process, or mimic approaches to modelling (see Burrough 1989, Walker and others 1986). Considerable research effort is being directed to improving the utility of deterministic process-based forest models (eg. McMurthrie and others 1989, Running and Coughlan 1989).

Empirical models are common in forest studies. These involve statistically quantifying the relations between observations made at a sample of field survey sites of (1) forest characteristics and (2) corresponding measurements or estimates of physical environmental variables. The goal is to predict the probability of occurrence of the biological phenomena in relation to values of specified environmental variables. These methods have the advantage of being applicable where process understanding is limited. Analyses can still proceed within a knowledge-based framework if independent variables are chosen that are hypothesised to be determinants of the biological phenomena under study. The choice of independent variables is one way in which theoretical notions of biotic-environmental relations can be incorporated into otherwise empirical analyses.

Given that some type of statistical analysis is adopted to quantify biotic-environment relations, a further stage in the analysis is then possible. If the distribution across the landscape of the independent variables used in the analysis is known, then a spatial prediction of the statistical function can be made. This approach has been adopted in the development of the MONOMAX algorithms, and procedures to link these with both biological site data and primary spatial data (see Bayes and Mackey 1991, Mackey and Bayes 1990). The MONOMAX package provides maximum likelihood estimates of a dependent variable as a function of one or two independent variables subject only to the assumption the function relating the variables is monotonic over at least part of the range of values. The algorithms are implemented using dynamic programming techniques, and are particularly suited to analyzing ranked data.

An example is given in Table 3. This is the result of a MONOMAX analysis based on tropical forest plot data recorded at 61 sites in the Wet Tropics of Queensland Australia (see Mackey in prep.). Here the probability of three classes of structural complexity is given as a function of (1) long-term mean monthly maximum temperature of the hottest month and (2) an index of position in catchment defined as

As/S (1)

where As is the specific catchment area (upslope accumulated area divided by the plot width), and w is the plot slope.

Class 1 represents sites with a simple structure, Class 2 have a medium level of complexity, while Class 3 represents the most structurally complex sites. Complex structures are most likely at maximum values of the temperature and catchment indices. Simple structures are most likely below a maximum temperature of 28 irrespective of the catchment index. As temperature increases however they are unlikely at increasing values of the catchment index. Thus the probability of the three classes can be predicted for any location in the Wet Tropics of Australia at which values for the maximum temperature and catchment indices are known.

Having the capacity to spatially predict the probability of a class of biological phenomena is extremely useful but further analysis is still required. Mackey ( in prep.) discusses field survey undertaken to compare predictions of tropical forest characteristics based on the above procedure. Where observed values are consistent with predicted then the hypothesis is judged to have been corroborated. However to fully implement scientific protocol for experimental ecology (such as suggested by Underwood 1990) requires the development and testing of a null-hypothesis. This means a suitable statistic is needed to test for significance. The nature of this statistic will vary, depending partly on the form of the predictive function.

In respect of the assumptions underlying the MONOMAX algorithms, a relevant statistical test for certain simple cases might be available in the Wilcoxon Two Sample Rank Sum Test (see Noether 1967). Suppose there are m observations of large leaves (L) and n observations of small leaves (S), and that these have been ranked according to some environmental criterion (eg. a thermal gradient), to produce a sequence such as

$S_1\ L_1\ S_2\ S_3\ L_2\ S_4\ L_3\ L_4\ S_5$ (2)

For any (SL) pair, say $S_i$ and $L_j$, the sign of (SL) is defined by

Sign $(S_iL_j) = 1$ if $S_i$ is to the left of $L_j$

Sign $(S_iL_j) = -1$ otherwise

The statistic W is the sum of the signs over all possible

$$W = \sum_{i=1}^{m} \sum_{j=1}^{n} Sign(S_i L_j)$$

(SL) pairs, that is

Under the null hypothesis that the S's and the L's in (2) are randomly sequenced, it can be shown that

$E(W) = 0$ (4)

and

$Var(W) = 1/3\, mn\, (m + n + 1)$ (5)

Provided m and n are not too small, the distribution of W can be assumed to be approximately normal, with mean and variance defined by (4) and (5). The usual statistical tests can be applied to this normal to detect a divergence from the null hypothesis. It is possible, though this has yet to be demonstrated, that the above test (or other tests from the theory of rank statistics) could be applied to the more complex one and two dimensional models developed in the MONOMAX algorithms.

**Table 3. Maximum likelihood probabilities for the mesophyll ratio of (a) a minimum temperature index and (b) a catchment index**

**Matrix class intervals:**

**(a) column variable is long-term mean monthly maximum temperature of the hottest month (degrees celsius)**
**(b) row variable is a catchment index (specific catchment area/slope)**

| Class intervals | column variable | row variable |
|---|---|---|
| 1 | 27 | -585 |
| 2 | 28 | 515 |
| 3 | 29 | 1000 |
| 4 | 30 | 1184 |
| 5 | 31 | 1485 |
| 6 | 32 | 1807 |
| 7 | 33 | 2252 |
| 8 | 34 | 2429 |
| 9 | | 3059 |
| 10 | | 3322 |
| 11 | | 5644 |
| 12 | | 5644 |

**Class (1) number of observations = 16**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 0.600 | 0.600 | 0.600 | 0.600 | 0.600 | 0.600 | 0.600 |
| 2 | 1.000 | 0.600 | 0.600 | 0.600 | 0.200 | 0.200 | 0.000 | 0.000 |
| 3 | 1.000 | 0.600 | 0.600 | 0.600 | 0.200 | 0.200 | 0.000 | 0.000 |
| 4 | 1.000 | 0.600 | 0.600 | 0.600 | 0.200 | 0.000 | 0.000 | 0.000 |
| 5 | 1.000 | 0.600 | 0.600 | 0.600 | 0.200 | 0.000 | 0.000 | 0.000 |
| 6 | 1.000 | 0.600 | 0.600 | 0.600 | 0.000 | 0.000 | 0.000 | 0.000 |
| 7 | 1.000 | 0.600 | 0.600 | 0.600 | 0.000 | 0.000 | 0.000 | 0.000 |
| 8 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 9 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 10 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 11 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 12 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 3 (cont'd)

**Class (2) number of observations = 31**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.000 | 0.400 | 0.400 | 0.400 | 0.400 | 0.400 | 0.400 | 0.400 |
| 2 | 0.000 | 0.400 | 0.400 | 0.400 | 0.709 | 0.709 | 0.999 | 0.250 |
| 3 | 0.000 | 0.400 | 0.400 | 0.400 | 0.709 | 0.709 | 0.999 | 0.250 |
| 4 | 0.000 | 0.400 | 0.400 | 0.400 | 0.709 | 0.909 | 0.909 | 0.250 |
| 5 | 0.000 | 0.323 | 0.323 | 0.323 | 0.709 | 0.909 | 0.909 | 0.250 |
| 6 | 0.000 | 0.323 | 0.323 | 0.323 | 0.909 | 0.909 | 0.909 | 0.250 |
| 7 | 0.000 | 0.323 | 0.323 | 0.323 | 0.909 | 0.909 | 0.909 | 0.250 |
| 8 | 0.000 | 0.923 | 0.923 | 0.923 | 0.909 | 0.588 | 0.500 | 0.250 |
| 9 | 0.000 | 0.923 | 0.923 | 0.923 | 0.909 | 0.588 | 0.500 | 0.250 |
| 10 | 0.000 | 0.588 | 0.588 | 0.588 | 0.588 | 0.588 | 0.250 | 0.250 |
| 11 | 0.000 | 0.588 | 0.588 | 0.588 | 0.588 | 0.588 | 0.250 | 0.000 |
| 12 | 0.000 | 0.588 | 0.588 | 0.588 | 0.588 | 0.588 | 0.250 | 0.000 |

**Class (3) number of observations = 14**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 2 | 0.000 | 0.000 | 0.000 | 0.000 | 0.091 | 0.091 | 0.091 | 0.750 |
| 3 | 0.000 | 0.000 | 0.000 | 0.000 | 0.091 | 0.091 | 0.091 | 0.750 |
| 4 | 0.000 | 0.000 | 0.000 | 0.000 | 0.091 | 0.091 | 0.091 | 0.750 |
| 5 | 0.000 | 0.077 | 0.077 | 0.077 | 0.091 | 0.091 | 0.091 | 0.750 |
| 6 | 0.000 | 0.077 | 0.077 | 0.077 | 0.091 | 0.091 | 0.091 | 0.750 |
| 7 | 0.000 | 0.077 | 0.077 | 0.077 | 0.091 | 0.091 | 0.091 | 0.750 |
| 8 | 0.000 | 0.077 | 0.077 | 0.077 | 0.091 | 0.412 | 0.500 | 0.750 |
| 9 | 0.000 | 0.077 | 0.077 | 0.077 | 0.091 | 0.412 | 0.500 | 0.750 |
| 10 | 0.000 | 0.412 | 0.412 | 0.412 | 0.412 | 0.412 | 0.750 | 0.750 |
| 11 | 0.000 | 0.412 | 0.412 | 0.412 | 0.412 | 0.412 | 0.750 | 1.000 |
| 12 | 0.000 | 0.412 | 0.412 | 0.412 | 0.412 | 0.412 | 0.750 | 1.000 |

## Conclusions

Resource managers are demanding more quantitative information on both ecological and timber values. Underpinning our approach to this problem is the development of a spatial data base for modelling the primary environmental regimes. While this is a significant undertaking, particular for large regions at a relatively fine scale, appropriate and feasible procedures are now available for many of the required attribute data and subsequent analyses (eg. the SPLINB, ANUDEM and MONOMAX procedures). Once achieved, the data base provides a useful tool for a wide range of forest-related issues. An example is the development of improved regional growth and yield models incorporating key climatic variables.

The Ontario project will operate at two scales: provincial-wide analysis at 1:250 000 and 1:10 000-20 000 for selected areas. Both are required for policy development and operational planning. The broader scale is often omitted but is essential for context and examining the total distributions of biological phenomena.

A stronger emphasis on primary data base development provides the key to making better use of biological site data. Field survey is expensive and it is essential that maximum information be extracted. Ontario is fortunate in having a number of significant collections of forest plot data, suitable for both ecological and timber analysis.

This paper has described some key aspects of database development for GIS analysis. Emphasis has been given to the need for quality control in all phases. This is particularly important for developing statistically reliable spatial predictions of biological phenomena. A greater emphasis on accounting for attribute data error and variability will ultimately assist planners and managers in dealing with the spatial complexity of the landscapes they are working in.

## References

Austin M.P., Cunningham R.B. and P.M. Fleming (1984). New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. Vegetatio 55:11-27.

Burrough P.A. (1989). Matching spatial data bases and quantitative models in land resource assessment. Soil Use and Management 5(1):3-16.

Bayes A.J. and Mackey (1991). Algorithms for monotonic functions and their applications to vegetation science. Ecological Modelling 56:135-159.

Fitzpatrick E.A. and H.A. Nix (1970). The climatic factor in Australian grassland ecology. In. Moore (Ed.) Australian Grasslands. Australian National University, Canberra, pp.3-26.

Hutchinson M.F. and R.J. Bischoff (1984). A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, NSW, Australia. Australian Meteorological Magazine 31:179-184.

Hutchinson M.F. (1987). Methods for generation of weather sequences. In. Bunting A.H. (Ed.) Agricultural Environments: Characterisation, classification and mapping. CAB International, Wallingford U.K., pp. 149-157.

Hutchinson M.F. (1989). A new procedure for gridding elevation and stream line data with automatic removal of pits. Journal of Hydrology 106:211-232.

Mackey B.G., Nix H.A., Hutchinson M.F., McMahon J.P. and P.M. Fleming (1988). Assessing representativeness of places for conservation and heritage listing. Environmental Management 12:501-514.

Mackey B.G. and A.J. Bayes (1990). A modelling framework for the spatial extension of ecological relations in vegetation studies. Mathematics and Computers in Simulation 32:225-29.

Mackey B.G. (in prep.). An environmental and spatial analysis of rainforest structure and physiognomy.

McMurthrie R.E., Landsberg J.J. and S. Linder (1989). Research priorities in field experiments on fast-growing tree plantations: implications of a mathematical production model. In. Pereira J.S. and J.J. Landsberg (Eds.) Biomass production by fast growing trees. Kluwer Academic Publishers, Dordrecht, pp.181-207.

Moore I.D., Grayson R.B. and A.R. Ladson (1991). Digital terrain modelling: a review of hydrological, geomorphological and biological applications. Hydrological Processes 5:3-30.

Murtha G.G. (1990). Soils of the Mossman - Cape Tribulation area. Divisional Report No.102, CSIRO Australia, Division of Soils.

Noether G.E. (1967). Elements of nonparametric statistics. John Wiley and Sons Inc., New York.

Running S.W. and J.C. Coughlan (1988). A general model of forest ecosystem processes for regional applications I. Hydrological balance, canopy gas exchange and primary production processes. Ecological Modelling 42:125-154.

Underwood A.J. (1990). Experiments in ecology and management : their logics, functions and interpretations. Australian Journal of Ecology 15(4):365-389.

Wahba G. and J. Wendelberger (1980). Some new mathematical methods for variational objective analysis using splines and cross-validation. Monthly Weather Review 108:1122-1143.

Walker J., Sharpe P.J., Penridge L.K. and Wu Hsin-i (1986). Competitive interactions between individuals of different size: the concept of ecological fields. Technical memorandum 86/11. CSIRO Division of Water and Land Resources, Canberra.

# B2 Visual impact modelling

**Viewscape analysis: a comprehensive methodology**

Charles Goodbrand, Reid, Collins and Associates,
Vancouver, BC, Canada

**Design and simulations of harvesting by visual quality objectives class**

Catherine Berris, Catherine Berris Associates Inc.,
Vancouver, BC, Canada

**Decision support tool for wood harvesting mangement in regard to landscape quality integrity**

Olivier Richer et al., Université Laval, Sainte Foy,
PQ, Canada

**GIS for visual impact assessment – a case study from the Three Sisters Resort development, Alberta**

David Buckley et al., TelTec GIS Division, Longmont,
CO, USA

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Viewscape analysis: a comprehensive methodology

**Charles Goodbrand**

Senior Consultant
Reid, Collins and Associates
A Division of H.A. Simons Ltd.
Suite 503, 425 Carrall Street
Vancouver, B.C. Canada V6B 2J6

## Abstract

*Reid, Collins and Associates, a division of H.A. Simons, provides a full range of engineering services to the forestry sector in both the Canadian and International marketplaces; Simons' core business is the design and commissioning of pulpmills. Ancillary services include strategic studies, feasibility studies, market analyses and environmental impact analyses. In this latter category of work Reid Collins had founda growing role for viewscape analysis as a necessary, but not sufficient, component of comprehensive environmental impact analysis.*

*In this paper we examine the objectives of viewscape analysis and discuss a comprehensive approach using Geographic Information Systems (GIS) to achieving those objectives including data and system procedures leading to the efficient generation of viewscapes. Some insight into when a particular technology is appropriate will result. We will also review some of the theoretical underpinnings required for this type of analysis and identify and discuss the area of most significant cost, the Digital Elevation Model.*

## Why viewscape analysis?

In any land use planning exercise, the public is increasingly involved and vocal in opposing any unnecessary alteration to the natural (or artificial) beauty of an existing viewscape. Approvals have been denied or unnecessarily delayed due to this concern. Public and representative special interest groups mistrust corporate and governmental intentions; reactive planning often contributes to this mistrust. When planners are inadequately prepared to provide acceptable answers addressing public concerns in a timely manner, the public assumes their interests are being compromised for the sake of expediency.

The British Columbia Ministry of Forests has been at the forefront in the development of an analytical methodology which stresses public involvement, public comprehension and technical feasibility. The Ministry has set out clear Visual Quality Objectives (VQO) for forest landscapes, criteriafor the classification of impact, and an understandable methodology, including standard presentation of the resulting analyses. This paper will review current computer methodologies for accomplishing these ends. All of the examples are drawn from existing GIS technologies and can be produced in a cost-effective manner.

Landscape design concerns itself with some very fundamental questions about the visibility of a cut-block from different vantage points. The current methodology posits a cutblock (designed by the forester in plan view) and assesses the impact of that design. The forester repeats the steps until an acceptable solution is derived. The methodology can be summarized as follows:

**(A) Design a cutblock, X, in plan view.**

1.0 From which areas is X visible? Identify all areas with viewscapes that are potentially affected by the development.

Method: manual drafting techniques, field experience, judgement calls.

Critique: may exclude viewpoints whichbecome contentious later, "missed views"; may include viewpoints which are not in fact visible, "red herring views".

2.0 Of these, which areas have viewscape values? These areas are project and study area dependent: scenic corridors, urban settlements, recreational areas.

Method: public forum, judgement calls, existing land use mapping.

Critique: the assignment of value is the contentious issue. "Missed views" from Step 1.0 will result in a great deal of acrimonious debate. "Red herring views" open unnecessay debate. New "red herring views" can be introduced because they have not been demonstrably excluded.

3.0 What parts of X are visible from these high value areas? Individually assess the extent of viewscape alteration for each of the intersecting areas (1.0 intersecting 2.0).

Method: manual drafting techniques, judgement calls.

Critique: the assumptions made are open to argument.

4.0 What does X look like from here? For each of the areas derived from 3.0, produce a simulated view.

Method: using computers, simulated views: draped views, video imaging, forest modelling (discussed below).

Critique: expensive to produce, based on assumptions derived in previous steps above.

In general, this methodology is dependent on the judgement of the designer. Acceptance by the public is reliant upon the trust invested in the designer. Since the acceptability of the design is ultimately grounded on a question of values, it is unlikely that any methodology can produce a perfect solution to this process; however, computers have been introduced into this process and they can be more effectively used in deriving solutions which are more objective.

The methodology using GIS, would be revised as follows:

**(B) Posit the maximum extent of potential cutblock, X, in plan view.**

1.0 From which areas is X visible? Identify all areas with viewscapes that are potentially affected by the development.

2.0 Of these, which areas have viewscape values? These areas are project and study area dependent: scenic corridors, urban settlements, recreational areas.

3.0 What parts of X are visible from these high value areas? Individually assess the extent of viewscape alteration for each of the intersecting areas (1.0 intersecting 2.0).

P.0 Design perspective views of cutblocks for each viewpoint identified in 3.0. Interactivelymove between views to generate an acceptable composite. Generate areas from resulting design for use in volume and economic feasibility studies.

4.0 What does X look like from here? For each of the areas derived from 3.0, produce a simulated view.

The use of GIS as an integral component of the process would provide the following benefits:

- a greater perception of objective results based on assumptions that can be clearly documented;
- an ability to eliminate "red herring views" definitively;
- the elimination of all "missing views" leading to more thorough analysis;
- the ability to provide simulations as part of the process rather than an additional expense;
- the ability to quickly calculate and make known the tradeoff between visual impact and economic benefit;
- cost-effectiveness will be achieved by eliminating the iterations of block design by executing cutblock design from plan view.

The remainder of the paper will discuss this methodology in more detail.

## Digital elevation models: the foundation

The theoretical foundation for viewscape analysis has been in-place since the mid-1970s. A quick review of that foundation, Digital Elevation Modelling (DEM), is appropriate focussing on the data and software requirements for DEM.

The input data structure for generation of DEMs are spot elevations, contours or regular rectangular gridded values; values must have northing, easting, and elevation (X,Y,Z) values. Data sources for such values are field survey notes, field computers (total stations, Global Positioning Systems (GPS), handheld computers, etc.), or existing manuscripts in either digital or analogue formats. Paper manuscripts are either digitized (digitally traced) or scanned (electronically traced using optical devices); photographic manuscripts are usually digitized using analytical stereo-plotters. There has been considerable successful research in processing remotely sensed satellite imagery in digital format to DEM formats.

The greatest part of the cost of DEM generation is data collection and verification. It is often the case that the data exists only in manuscript form, or, if already digital form, is not at a resolution suitable for the task at hand. Our work in site preparation or cutblock design frequently calls for data at 1:2000 or 1:5000 scales when available digital sources are at 1:20 000 or higher scales. We often need to re-survey and create appropriate DEM bases. The analyst must be wary of using data sets with inappropriate abstractness: beautiful but misleading products are too easily generated.

Software required to process these input structures usually provides the capability of creating four digital representations of the DEM from any one input source: spot elevations, contours, Triangular Irregular Network (TIN), or regular grid. These digital representations are also products of DEM in that they can be generated in manuscript or tabular format. In tabular digital format the DEM can be used as input to other programs for volumetric computations, haul distances, etc.

Viewscape analysis is one of those potential uses. This type of analysis is dependent upon a base DEM at the appropriate level of abstraction. The manipulations performed on the DEM fall into two general categories: object insertion and draping. Object insertion requires surface or solid models of objects with known elevations in digital format. H.A. Simons often creates these models as part of the engineering design process in a Computer Assisted Design and Drafting (CADD) environment. Draping is the process of superimposing digitally encoded planimetric areas (polygons) onto the DEM. Figure 1 illustrates these concepts.

A viable viewscape analysis package is able to build this base DEM, provide a simple user interface, and must make allowances for data exchange between different sources of data.

A variety of end products can be created from this theoretical base; however, the validity of the analysis derived from all other steps of the viewscape methodology are dependent upon the reliability of the DEM.

## View selection

When utilizing GIS software with a DEM in Steps 1.0, 2.0, and 3.0 of the viewscape methodology, polygon analysis is the most useful approach. Figure 2 illustrates this process.

We assume a DEM has been created and that the GIS can perform simple polygon analysis on the DEM. The GIS software should allow for the definition of multiple viewpoints and be capable of creating visibility polygons.

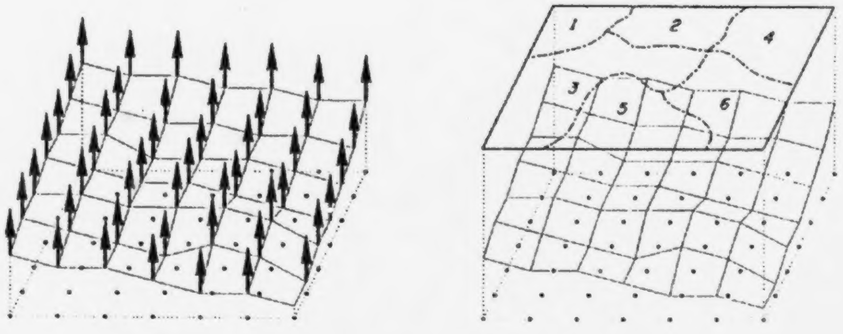


**Figure 1:** *Object Insertion vs Surface Draping*

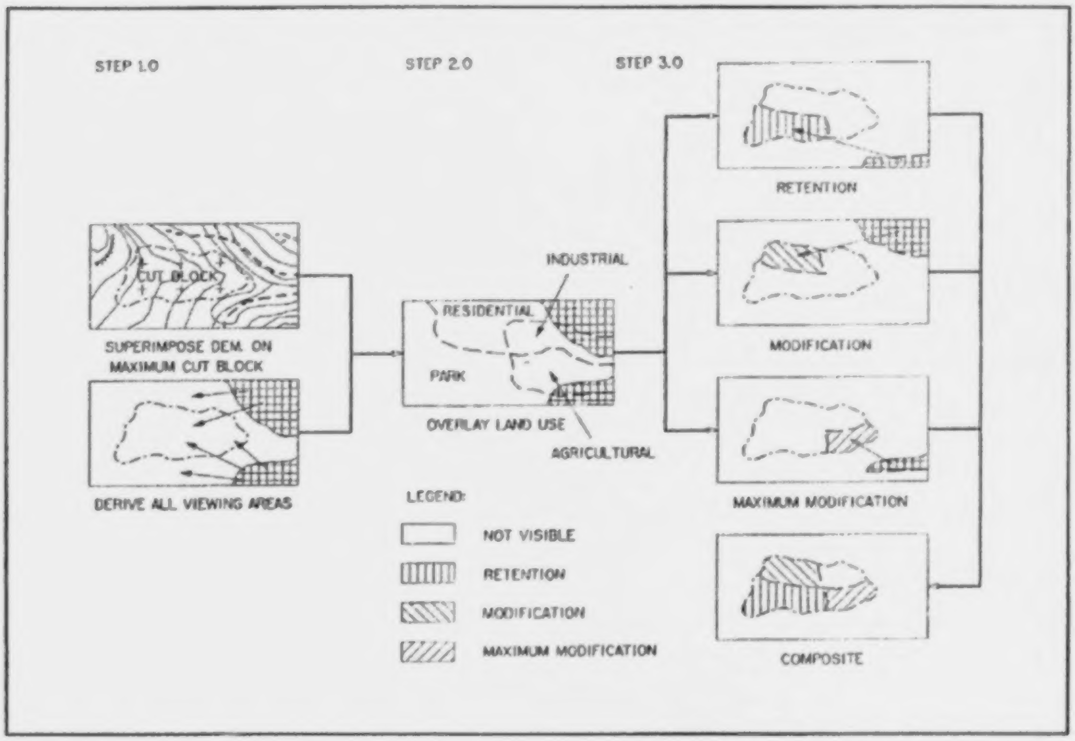

STEP 1.0
STEP 2.0
STEP 3.0
CUT BLOCK
SUPERIMPOSE DEM. ON MAXIMUM CUT BLOCK
DERIVE ALL VIEWING AREAS
INDUSTRIAL
RESIDENTIAL
PARK
OVERLAY LAND USE
AGRICULTURAL
RETENTION
MODIFICATION
MAXIMUM MODIFICATION
COMPOSITE
LEGEND:
NOT VISIBLE
RETENTION
MODIFICATION
MAXIMUM MODIFICATION

**Figure 2:** *Polygon Analysis*

Expanding Step 1.0 of the viewscape methodology:

1.1 define a closed perimeter for the development area;
1.2 define representative viewpoints within this perimeter;
1.3 generate the viewable areas;
1.4 interpret the result as areas from which some or all of the development is viewable.

Note that unless special steps are taken, this analysis will provide theoretical view areas, that is, view areas without regard to ground cover or local variations in terrain smaller than the DEM resolution.

Step 2.0 applies polygon analysis by overlaying these polygons and others and deriving intersecting polygons which have composite attributes corresponding to Visual Quality Objectives: preservation, retention, partial retention, modification, maximum modification, or excessive modification. Overlay analysis assumes that themes of interest — scenic corridors, urban settlements, recreational areas and so on — have been digitally encoded as closed perimeters with a visual impact ranking. As an example of a detailed site-specific analysis, it may be predetermined that all Provincial Park viewscapes should be retained, that residential viewscapes can be modified, and that land classed as agriculture can be planned for maximum modification. A query to derive intersection areas of the resultant viewable areas and the impact areas will yield those areas where more detailed study may be required.

Where more general analysis is to be undertaken, visual impact information can be derived from the British Columbia Ministry of Forests' Recreational Overlay as follows (see Visual Absorption Capability chart).

An initial assessment of the severity of the problem (Step 3.0) can be achieved through a repetition of individual queries as in Step 1.0 modified to analyze each of the impact areas identified in Step 2.0. An individual analysis for each high impact area will:

2.1 define representative viewpoints within the perimeter;
2.2 generate the viewable areas of the potential cutblock;
2.3 interpret the result as areas of the development area viewable from the high impact area and assign the VQO accordingly.

**Visual Absorption Capability**

| SENSITIVITY | None | Low | Medium | High |
|---|---|---|---|---|
| High | $V^1$ | $V$[illegible] | $V^5$ | $V^6$ |
| Medium | $V^2$ | $V^7$ | $V^8$ | $V^9$ |
| Low | $V^3$ | $V^3$ | $V^3$ | $V^3$ |

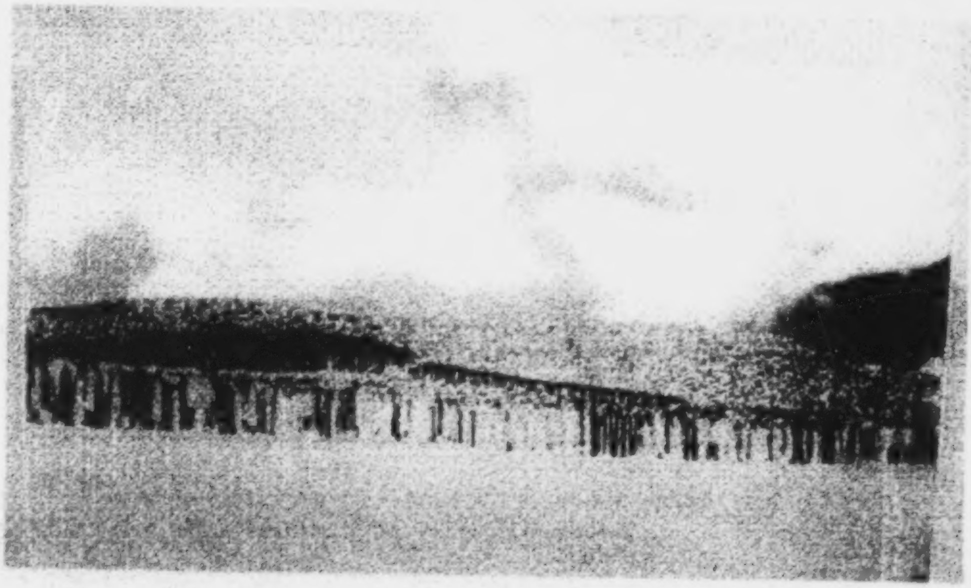


**Figure 3:** *Modelling the Forest*

This analysis should provide the planner with a thorough list of potential problems selected on the basis of criteria agreed to by all parties; the resultant analyses will be independently generated based on the givens and therefore will be perceived as unbiased and objective; the process is thoroughly self-documented. With this documentation, the planner should be able to answer public concerns about areas not affected by the development, plan mitigation for those areas which are viewable, and prepare more detailed analyses on areas of highest impact.

## Step P.O: perspective design

The placement of cutblocks for forest harvesting is of great importance to Woodlands Management groups. The example provided in Figure 3 illustrates the degree of realism that can be simulated by viewscape software. This example was produced using Intergraph Systems Ltd. DEM software. The individual trees are inserted onto an INTERGRAPH surface model, with shape and height derived from inventoried species composition and age height class. The DEM values were generated from Government of British Columbia maintained, digitally encoded, 1:20 000 Forest Cover Maps.

This form of rendered viewscape is of great value for ad hoc queries. It combines hidden line removal software, familiar to CADD software users and DEM software. Single viewpoint requests can be created in a matter of minutes. Production time for the finished product will vary with the complexity of the DEM, density of the trees, and speed of the computer.

Perhaps of more interest to the forester is that this viewscape was designed from the perspective view.

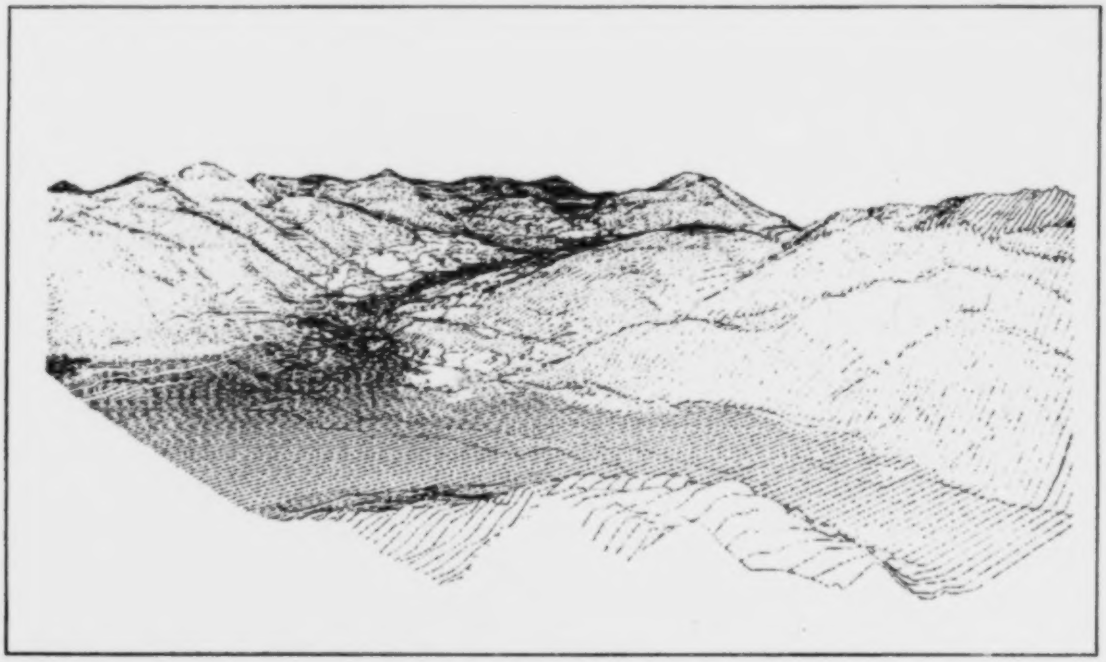


Figure 4: *Draped Views*

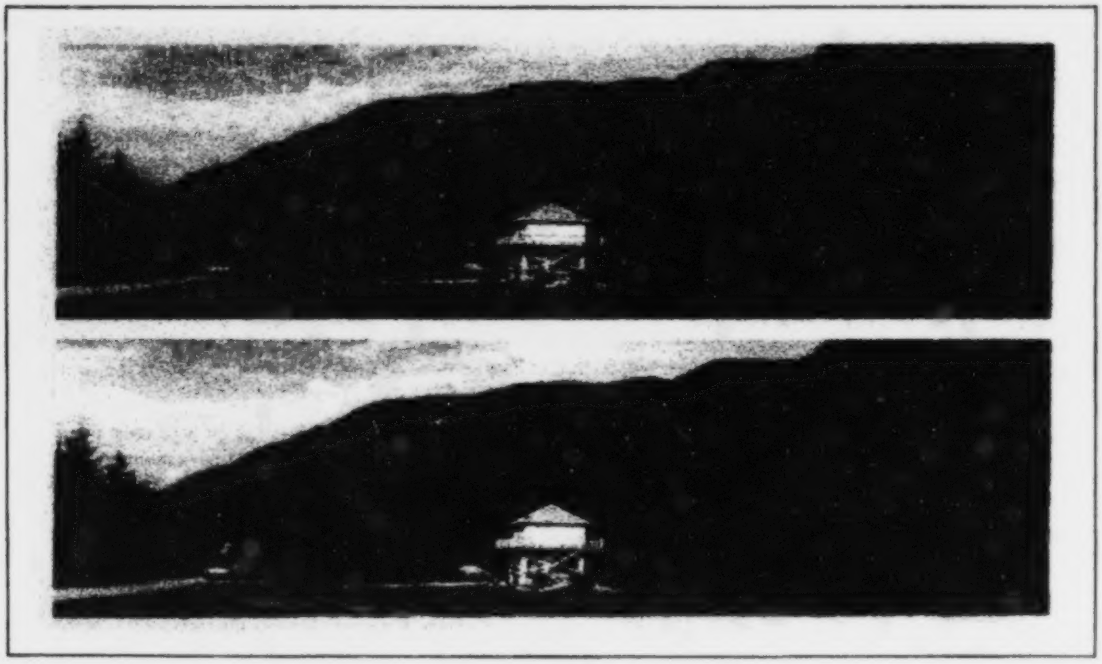


Figure 5: *Video Imaging*

The cost of the view production is related to the digital base data, the DEM, and forest cover information. Redesign of the cutblock can be accomplished in a matter of hours. The planimetric information of the cutblock is a by-product of the process, hence area of cut and economic calculations can be accomplished immediately following the design.

## Viewscape presentations

As stated, draping is the process whereby a planimetric area is overlain on a DEM in such a way that the planimetric area takes on the terrain of the DEM. The technique is commonly used with colour theming to produce viewscapes from a distant perspective. Figure 4 was created on Digital Resource's TERRASOFT system.

Draped views are often augmented with a relative Z above ground which is imposed on the planimetric data. This augmented Z allows the simulation of homogenous natural land coverage: forests, grasslands, etc.

Video imaging is often utilized in a variation on draping used where highly realistic viewscapes are required. It entails photographic capture of a "before" view from a known perspective. The image is captured in a frame buffer and draped on a DEM aligned to that viewpoint. Other photography, of a clear cut, for example, is digitally "cut in" to the picture. A new "picture" is produced and either printed by a high resolution colour printer or photographed from the computer monitor. The same DEM used for the the rest of the methodology is used for the creation of the video image; the perspective view information is used to control the placement of the "cut in" information. Figure 5 illustrates this principle.

## Conclusion

As the public becomes increasingly concerned with the degradation of the tangible environment and the intangible aesthetic environment, it will become critical to prompt project start-up to have anticipated these public concerns. Viewscape analysis is a technique which will allow planners to anticipate some of the intangible aesthetic concerns. The planner with a well conceived methodology will be able to take full advantage of current hardware and software technology.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Design and simulations of harvesting by VQO class

**Catherine R. Berris**[1]

Catherine Berris Associates Inc.
250 - 1508 West 2nd Avenue
Vancouver, B.C. V6J 1H2
Tel: 736-6336
Fax: 736-2338

### Abstract

*A project was undertaken to study the interrelationships between harvest designs and VQO classes, and to assess the capabilities of simulation technology. Four separate harvesting designs were developed for each VQO class, two with one cutblock and two with multiple cutblocks. The cutblocks were measured in plan and perspective to determine the area harvested as a percentage of the landscape management unit. Designs were refined to ensure that they fit within the VQO classes. The result provides an interesting comparison of harvesting options, and demonstrates the capabilities of the methods used. Tools used were GIS, digital terrain modelling (DTM), and computer imaging (CI) on an IBM-compatible system. The project was sponsored by the Ministry of Forests.*

## Introduction

*Background:* In making forest management decisions, a fine balance is often required to achieve acceptable visual quality while simultaneously accomplishing other resource objectives. Simulating the visual appearance of planned cutblocks and road development is a critical factor in this process. The public is becoming more sophisticated and wants to be able to see the impacts of proposed alterations. Simulation is also the most effective tool available for forest landscape design.

This project was initiated by the Ministry of Forests Recreation Branch. There were two primary objectives:

- to contribute information to the process of quantifying the amount of alteration acceptable within the various Visual Quality Objective (VQO) classes,
- to assess the current capabilities of computer imaging (CI) combined with DTM in simulating proposed alterations.

## Method

One viewpoint was selected for the project and a landscape unit within that view was identified for study. The viewpoint, overlooking Mount Seven near Golden, B.C., was selected by the Ministry of Forests because of interest in the area for landscape management study, and because good information was available. Photographs from that viewpoint were taken with a 50 mm. lens and a level camera at two different times on a sunny day in fall.

The two photographs were scanned into the computer, and imported to the imaging software TIPS. Because of the low angle of the sun due to the season, one photograph was in shadow at the top, and the other had extensive shadow at the bottom. The two photographs were merged into one, combining the sunny portions of each.

A 1:20 000 map of the area (B.C. TRIM series) was downloaded into the GIS, Terrasoft, and a digital terrain model of the study area at 65 metre resolution was generated. Using the known viewpoint location, a computer perspective from that point was produced, and it was aligned with the photograph to ensure an accurate match. A visibility map from the viewpoint was generated and plotted, and the landscape unit boundary was drawn in plan, plotted in perspective and refined until it matched the landscape unit boundary as seen in the photograph. This step was important to ensure that cutblocks would not be placed in non-visible areas beyond the landscape unit boundary, thus skewing the results.

The next task involved developing designs of cutblocks to meet the VQO classes: Retention (R), Partial Retention (PR), Modification (M), and Maximum Modification (MM). Cutblocks were laid out in plan, digitized into the computer, and drawn in perspective by the computer. They were modified in accordance with landscape design principles to derive a palette of two to six cutblocks for each VQO, more being required for R and PR cutblocks.

The cutblocks were then combined together into sixteen scenarios, 4 for each of the above VQO's, with two of the four showing single cutblocks, and the other two meeting the VQO with multiple cutblock designs. Since the situation was considered completely hypothetical, little attention was paid to possible road layouts, and there was no minimum cutblock size, although no cuts were designed smaller than 6 hectares.

Once sixteen acceptable computer perspectives were produced, the cutblock areas were measured as a percentage of the landscape unit in plan and in perspective.

The final task involves generation of images which simulate the cutblocks as they will look when clearcut. Since the existing view includes one cutblock which is greening up, that cutblock will be filled in with mature trees to simulate the Preservation (P) VQO. The other simulations will be produced by inserting clearcuts from photographs of similar areas into the scene. At the time of writing this paper, the simulations have yet to be produced. They will be completed and presented as slides at the GIS 92 conference.

## Discussion of results

Figure 1 shows eight of the sixteen computer perspectives which illustrate the cutblock layouts. The landscape unit boundary is also indicated on the perspectives. At the time of writing this paper, these designs are considered drafts only and they may be revised.

The cutblock area percentages are presented in Table 1. The table also compares these results with Ministry of Forests draft guidelines, which are intended to be measured either in plan or perspective. In measuring cutblock alterations in perspective, the question arose as to whether the cutblock area should include only the ground surface visible within the cutblock, or whether it should also include the face of the trees behind the cutblock. In some Retention cuts, only the trees behind are visible. Both measurements have been taken and the implications will be studied further. For the benefit of this paper, the latter measurement including the trees behind has been used, since seeing the trees represents an alteration from seeing a solid mass of tree-tops only.

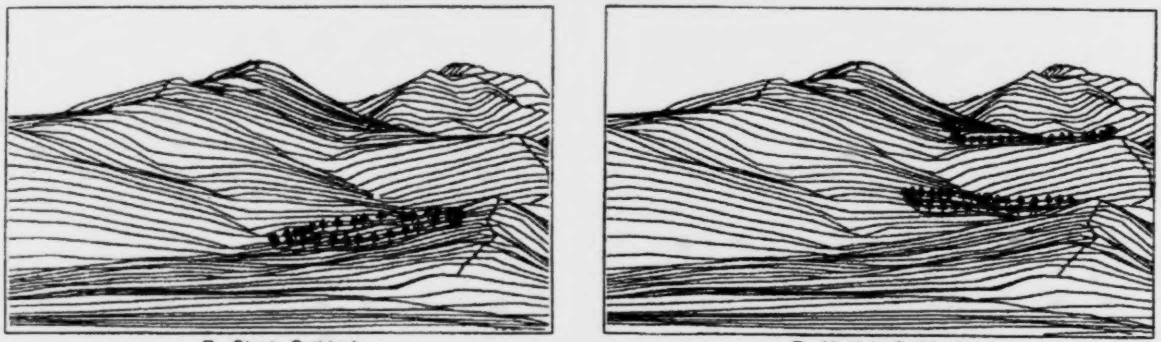
The following observations have been made based on the preliminary analysis:

1. Guidelines related to percentage of area cut in plan view need to take distance into account to be meaningful. The landscape unit studied extended from 2 to 5 kilometres from the viewpoint. While this would all be considered middleground, cutblocks towards the back of the unit could be significantly larger than cuts in the front portion while still meeting the same VQO.
2. Guidelines related to percentage of area cut need to be different as measured in plan from those measured in perspective to be consistent. The plan measurements from this exercise resembled the MOF guidelines though in this project, the numbers were generally lower. The areas as measured in perspective were significantly lower than the MOF guidelines.
3. The average cutblock area as measured in plan was generally higher for the scenarios with multiple cutblocks. This implies that it is possible to cut a larger area within the same VQO by distributing the cut area among two or more cutblocks.
4. The test study area has a moderate Visual Absorption Capability (VAC). While it is fairly steep generally and oriented directly towards the viewpoint, the topography does include several significant benches. In landscapes with lower or higher VAC's, the sizes of cutblocks would likely need to be reduced, or could be increased accordingly.

Further experimentation and investigation will be undertaken to further refine the analysis.

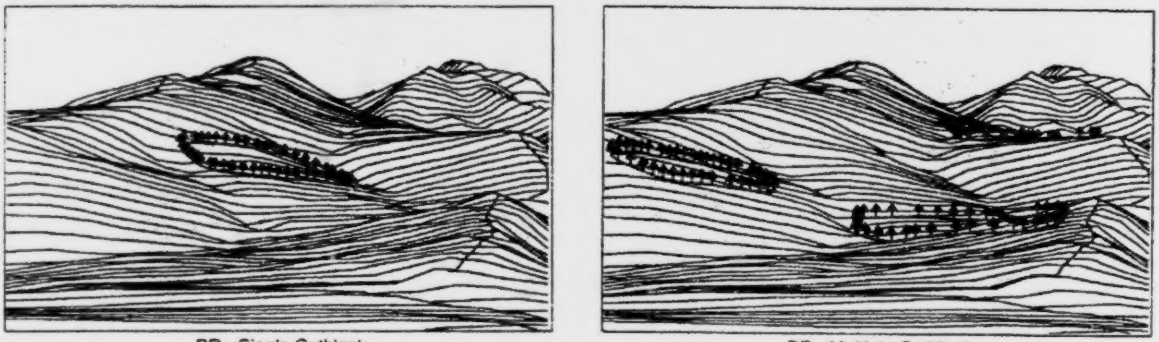
Figure 2 shows a sample simulation produced using imaging. It should be noted that these are intended for reproduction in colour, and that the clarity is compromised by the printing method used. The ability of the colour simulations to be used for different purposes once they are all completed will need to be assessed. Early indications are that they are certainly useful for public presentation and as design tools. Whether they

**Figure 1: Sample Computer Perspectives**



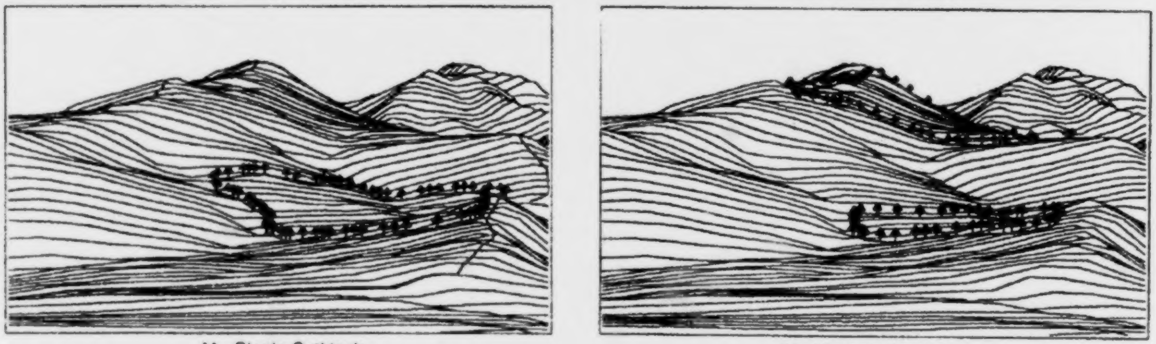
R - Single Cutblock

R - Multiple Cutblocks



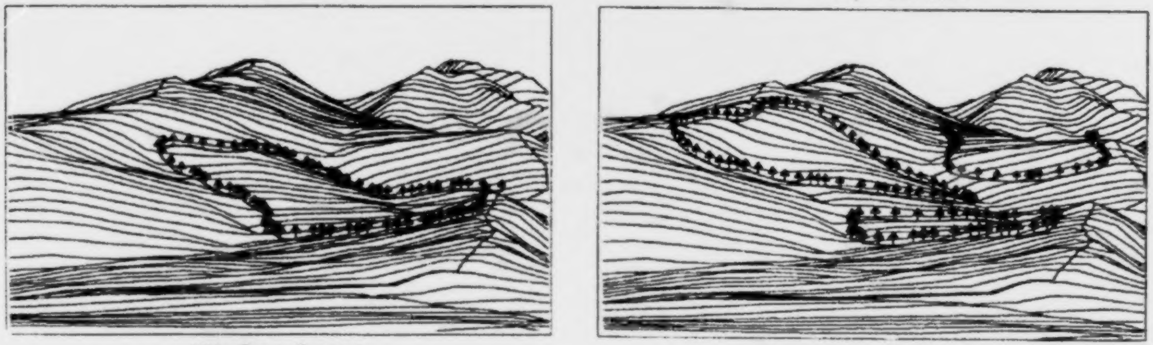
PR - Single Cutblock

PR - Multiple Cutblocks



M - Single Cutblock

M - Multiple Cutblocks



MM - Single Cutblock

MM - Multiple Cutblocks

Figure 2: Cutback Simulation

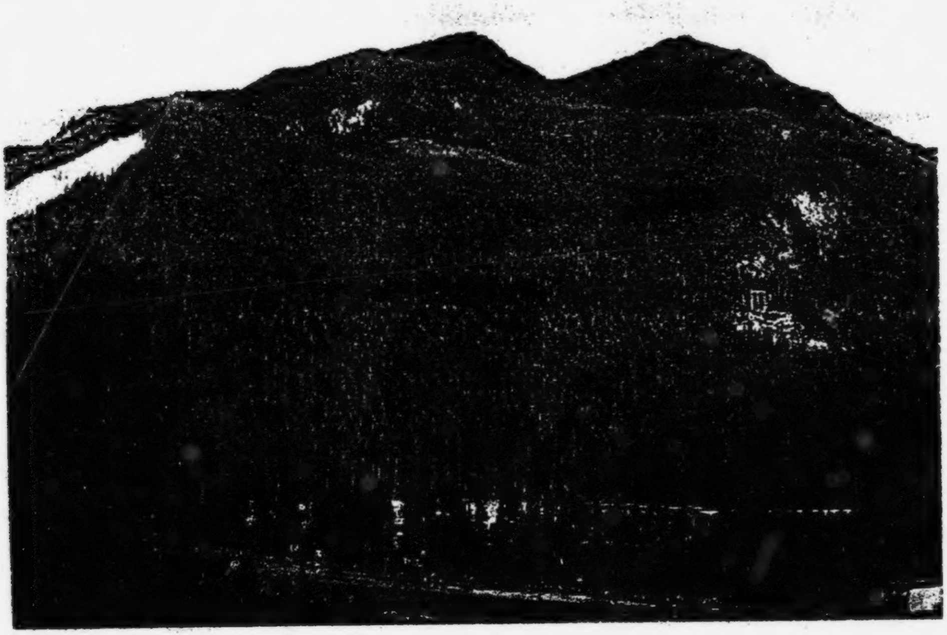


Table 1: Cutblock Area Percentages

| VQO Class | % Cutblock In Plan | % Cutblock in Perspective | MOF Guideline |
|---|---|---|---|
| **R** - single cutblock | 3.3, 5.6 | 2.4, 2.6 | 0 - 5% |
| **R** - multiple cutblocks | 2.5, 9.9 | 2.4, 2.6 | |
| **PR** - single cutblock | 3.0, 5.0 | 3.1, 3.4 | 6 - 15% |
| **PR** - multiple cutblocks | 5.8, 10.7 | 5.3, 6.2 | |
| **M** - single cutblock | 15.3, 15.9 | 6.4, 9.9 | 16 - 30% |
| **M** - multiple cutblocks | 18.8, 14.5 | 8.6, 11.4 | |
| **MM** - single cutblock | 9.5, 19.2 | 11.8, 13.5 | 31 - 50% |
| **MM** - multiple cutblocks | 24.2, 33.4 | 18.0, 20.7 | |

are clear enough to be used in photo-questionnaire perception studies will need to be determined at a later date.

One limitation of this project was the fact that it did not investigate alternate logging methods such as selective logging. That type of analysis will be important in the future as more alternative methods of logging are implemented.

The methods used in this project were sufficient to accomplish the objectives. The speed and potential results of the work, however, are hampered by current limitations of GIS technology. Arriving at suitable cutblock designs currently involves a "trial and error" approach, and the need to draw cutblocks in plan then to view them in perspective. What would greatly enhance this process is a "live" DTM, which would enable the user to identify a cutblock in perspective and have the computer plot the same area in plan.

## Bibliography

Ministry of Forests. *Forest Landscape Handbook.* Victoria: Province of British Columbia, 1981.

Ministry of Forests. *Interim Forest Landscape Management Guidelines for the Vancouver Forest Region.* Victoria: Province of British Columbia, 1990.

---

[1] *Special thanks to Kamill Apt of Ministry of Forests who initiated the project, Larry Price of Ministry of Forests who took the photographs, and Brian Pearson of Catherine Berris Associates Inc. who was the computer analyst.*

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Decision support tool for wood harvesting management in regard to landscape quality integrity

# Outil d'aide à la planification des récoltes forestières dans une perspective de préservation de la qualité du paysage

**Olivier Richer, ing. f.**

Alberta-Pacific Forest Industries Inc.
P.O. Box 1313
Edmonton, Alberta
Canada, T5J 2M8

**Dr Jean-Jacques Chevallier**

Centre de recherche en géomatique
Université Laval, Pavillon Casault
Sainte-Foy (Québec)
Canada, G1K 7P4

## Abstract

*In forestry, wood harvesting must respect the visual environment of sites devoted to recreation and preservation. To satisfy both the forest industry needs and the landscape quality requirements, a complex decision-making process must be followed: one must choose the appropriate forest stands and the logging periods to execute the logging operations. The actual tools are not sufficient and the existing GISs must evolve toward a true SDSS.*

*Starting from that initial acknowledgement, this paper describes a decision-making method based on simulation and identifies the new characteristics that a GIS must have in order to meet the planning needs of an organisation.*

## Résumé

*En foresterie, la récolte des bois doit respecter l'encadrement visuel des sites voués à la récréation et la conservation. Satisfaire à la fois les besoins de l'industrie forestière et les exigences de qualité de paysage nécessite un processus de prise de décision complexe: il s'agit de choisir adéquatement les peuplements forestiers et les périodes de coupes où les interventions seront réalisées sur le territoire. Les fonctionnalités actuelles ne sont pas suffisantes et il faudrait faire évoluer les GIS existants vers de véritables SDSS.*

*A partir de cette réflexion initiale, cet article décrit une méthode de prise de décision basée sur la simulation et identifie les nouvelles caractéristiques que devront avoir les GIS pour répondre aux besoins de la planification.*

## Introduction

The diversity of forest resources and the multiple needs of users force us to view different axes of development. Consequently, strategic objectives must be established. The planning process, which results from the chosen strategies, contains several options. The decision-maker will choose the best one according to his *assessment criteria*.

In this context, the role of the existing SIS is most often limited in the replication of good cartographic functions used by a competent specialist; the improvements lay essentially within the *speed* and the *repetition capability of the analitical functions*, the capacity (under certain conditions) to *master and access easily large volumes of data* and the *representation flexibility* .

Nothing in the actual software exists specifically to assist a decision-maker in his definition of objectives, in their translation into different forms of actions, in the assessment of their advantages and drawbacks, and in the selection of the optimal solution.

## Actual SIS application in forestry

### 2.1 Goal of SIS

The conventional SIS's have been developed to manage data that describe the territory, namely the forest territory. These data may be different in nature in regards to there source: maps, aerial photographs, data bases, satellite images. The model we have built of the forest shows us the way to collect the data. Once the inventory is done, the state of the forest can be described.

Therefore, one may acquire better knowledge of the forest territory from the data analysis capabilities of the SIS. This knowledge will help foresters to make decisions concerning forest exploitation; this use of SIS is mostly used at an operational level to make stand-alone decisions.

However, it would be desirable that an SIS assist decision-making at a higher level and thus allow the translation of the general objectives of an organization in detailed operations.

### 2.2 Description of forest territory

A forest territory can be seen as a collection of different entities serving a defined purpose. These entities will be refered to as **forest objects**. For example, forest stands, streams and roads are such objects each of them having particular *characteristics*. In regards to the sought-after objectives, the description of the forest will be more or less complex. One could consider the objects individually or collectively, within a specific group or within a population.

#### *2.2.1 Characteristics of a forest object*

A *forest object* could be described individually by its **descriptive properties** and its **spatial components**. The *descriptive properties* represent the characteristics which the forester gives to an object to describe it as it exists in the real world and in regards to his goals. The *spatial components* allow the description of a forest object in spatial dimensions according to its position, its orientation, its form and its size (Chevallier, Bédard 1991).

#### *2.2.2 Relationship between forest objects*

A forest object could also be located in relation to another object or a group of objects. The **relationship,** descriptive or spatial, that the object has with the others are then described (Champoux 1991). For example, a forest stand may be located near to a road, on fertile soil or partly within a green band; a river is composed of ordered sections.

#### *2.2.3 Distribution of a population*

A *set of objects* having the similar characteristics may form a **population**. In addition to the characteristics of each objects, a population possesses attributes that define its spatial distribution in the forest territory (Bilodeau 1991).

#### *2.2.4 State of forest territory*

Taking into account the particularities mentioned above, the **state of the forest territory** is determined by these three aspects:

- the set of the descriptive properties and spatial components of the individual objects that compose the forest;
- the set of descriptive and spatial relations between any two objects;
- the set of spatial distributions of the global set of objects or the populations.

It is important to note that this definition does not include the temporal aspect. Only the instantanous state of the forest is considered without any regards to present, past or future.

### 2.3 Historical data *vs* extrapolation

It is possible to keep in an SIS historical data which describe the past evolution of the forest object, for example the silvicultural operations performed on a given site. The techniques used to manage this historical data are varied but it is not our concern to analyze them in detail.

Based on historical data, the future can be "predicted" in certain cases. In other words, the evolution of certain phenomena described in an SIS can be extrapolated over a period of time (Lowell 1989, 1991). This can be seen when:

- the nature of the phenomena allows it (e.g. the growth of forest stand);
- the influencing conditions of this evolution will not change significantly;
- the duration of extrapolation is relatively limited.

These functions of extrapolation are not usually included in an SIS, but they are more often independant applications using inventory data of an SIS.

## About "decision making"

### 3.1 To decide in land management

In land management, **to decide** means "to choose the most appropriate action to undertake on the territory in order to realize a given objective". Thus, the process of decision includes three basic components:

- the *territory* on which the desired action will produce its effect;
- the *action* itself;
- the *decision-maker* who chooses the desired action.

In order to choose, the decision-maker must assess the impact of the desired action or group of actions. The assessment focuses on:

- the *action* itself;
- the *effect* resulting from the action.

The assessment of the *action* must make sure that the action will meet the established objectives and that the action will be carried out according to the rules of the art.

The assessment of the *effect* of one action or a group of actions on the land takes into consideration other factors such as:

(a) the *global contribution* of a set of actions towards the *objectives*;

(b) the *impact* of one or several actions on the *environment* of the objects.

In this last case, the assessment is in fact *the comparaison of the territory state before and after the action.*

Thus, the assessment of an annual forest operation plan can be performed for example:

- in regards to its production in volume or in cost (goal *a* mentioned above);
- in regards to its impact on the soil erosion or in the quality of the landscape (goal *b* mentioned above).

In the perspective of a complex planning process involving several steps, the assessment should be spread out over a certain period of time in order to take into consideration the expression of phenomena at differnt moments resulting from the desired action. Thus, the assessment will be done:

- at ***short term***, just after the action is completed (resulting state);
- at ***mid term***, after the direct effects of the action are toned down;
- at ***long term***, in order to measure the "permanent" effects of the action.

The choice of the "adequate action" rests on the decision-making process comprised of three parts:

- the *simulation* of the action;
- the *extrapolation* of the future states of territory at short, mid and long term intervales;
- the *global assessment* of the impact of action at short, mid and long term intervales.

The SIS, on which the decision-making process is based, must allow the management of data needed for the selection of the optimal action. In order to perform an impact assessment, the decision-maker must be able to have access to all the territory states. In other words, to contribute best in decision-making, a SIS must have the capability to keep and manage the several states resulting from the simulated development of the territory.

In the case when a state may result from a series of actions spread out in time, the decision-maker should be able to build or retrace the path leading to the final state. Consequently, to facilitate the creation of various forms, the system structure will then allow the modification of an action plan from a stage without having to redo it all from the beginning. So it is important to memorize all intermediate transactions i.e. to keep trace of transaction which have led the assessed states.

### 3.2 Changes of a forest state

#### *3.2.1 Modification of an object in time*

Until now, we have considered in an SIS only the instantanous state of an object, the one which exists at a given time (except for the historical data discussed in section 2.3 above). When considering time dimension, one can see the change of state of an object, i.e. the alteration in time of the value of its descriptive or spatial characteristics (Gagnon 1992).

On an axis of progression, where states and changes alternately succeed, we define a **link of progression** as a set including (*cf* figure 1):

- the *initial state* of an object arbitrarely defined;
- the *change* which modifies the object;
- the *final state* resulting from the change.

In a complex planning process, there will be a succession of links where the final state of a given link corresponds to the initial state of the following link.

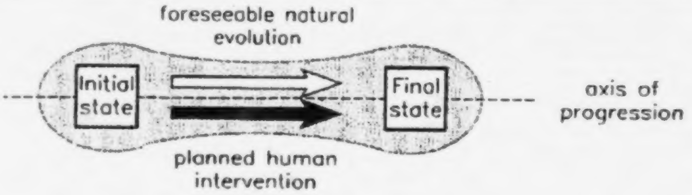


Figure 1 – Link of progression

When an inventory of the territory could be represented as an instant picture of a piece of land, the axis of progression may be seen as an ordered set of pictures taken at different moments. In the same perspective when one keeps historical data in an SIS, he memorizes a succession of states of the territory.

*3.2.2 Factors of change*

In the context of land development, men intervene in the environment by planning their actions in order to achieve the developmental goals. In this perspective, an object could be modified by two types of factors:

- the **internal factors** of the planning process which include the desired actions;
- the **external factors** of the planning process.

The *internal factors* include all planned human interventions, e.g.: cut and fill works, wood harvesting and road construction.

The *external factors* are grouped in three classes: the human interventions, which have not been expected within a planning proces; the spontaneous evolution parameters, which correspond to the natural evolution as vegetal growth; and the unforeseeable natural events as forest fires and floods.

In this paper, we take into account only the planned human interventions and the foreseeable natural evolution parameters. We will consider that a state could be (a) the result of the foreseeable natural evolution only or (b) the result of both the foreseeable natural evolution and the planned human interventions. We will not talk neither about the probabilistic study of the localisation of unforeseeable natural events, such as insect epidemics, nor the effect of this type of event on operation planning.

*3.2.3 Action, group of actions, program of actions*

The planned human interventions are graded in a hierarchical order. On the elementary level, the **action** is the basic component: it is described by its explicited effect on *one* territory object. For example, an action could be the harvesting of a particular forest stand or the construction of a specific road.

At a more general level, we define a **group of actions** as a *set of closely related actions being done in a given period of time*. In the planning perspective, the group of actions acts as a whole on the state of a set of objects. The actions, which compose a group of actions, are considered *simultaneous*. For example, an annual operation plan, which is a set of individual operations, changes the state of a set of forest stands.

When planning period is considered, the set of the groups of actions is called a **program**. A program is an *ordered series of the groups of actions*. Every group of actions corresponds to a step of the program (*cf* figure 2). A five year operation plan, for example, could be considered as a program.

### 3.3 Planning the changes

*3.3.1 Series of potential axes of progression*

From a given initial state of the territory, if one possible action is considered, two options can be seen:

- there will be an action;
- no action is planned.

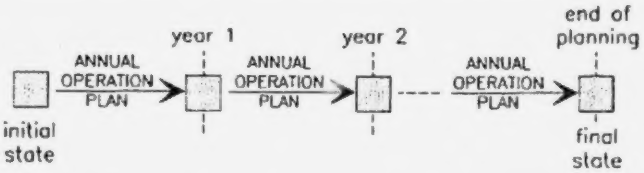


Figure 2 – Program of actions

Thus, the decision-maker must choose between *two divergent links of progression* with their related final states. Beginning with these states, if we proceed the planning to the following step, the same situation will occur and there will be four resulting final states.

From a general point of view, a set of **n** possible actions at a given step generates **n + 1** *divergent links of progression* (*cf* figure 3). Each series of links leading to a final state constitutes a particular program of actions and generate a distinct potential axis of progression. The set of all these potential programs of actions defines the *solution set* of a given planning problem.

A particular case needs to be considered. It is possible that the final state of a given project is already fixed. Even in this case, every possible evolution path that may lead to the desired final state will corresponds to a distinct program of actions.

*3.3.2 Modification assessment*

The planning process is based on actions. However, the decision-maker's choice is made according to the results of the assessment of the final states. Wood harvesting programs are planned in regards to the results of these harvestings. Thus, the purpose of the planning is not as much in the actions themselves as in the impact they will produce (*cf* figure 4).

The assessment of the resulting states is then indispensable in decision-making. The assessment process will use different criteria based on the impact that one would like to measure. The basis of assessment may have different forms, such as thematic images, reports, maps, etc., that are called **views.** These views are *partial representations of the inventory data* and they are produced for a particular purpose.

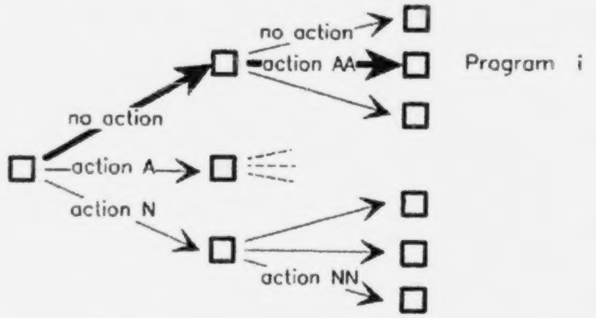

no action
action AA
Program i
no action
action A
action N
action NN

Figure 3 – Different axes of potential progression

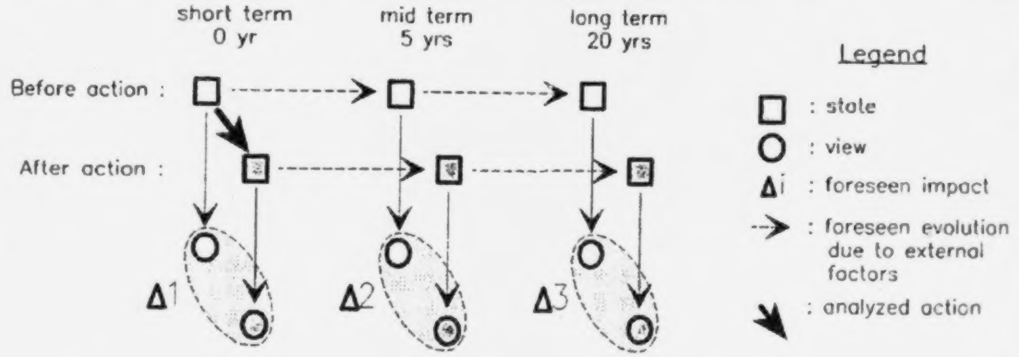

short term
0 yr
mid term
5 yrs
long term
20 yrs
Before action :
After action :
Δ1
Δ2
Δ3
Legend
□ : state
○ : view
Δi : foreseen impact
--> : foreseen evolution due to external factors
: analyzed action

Figure 4 – Impact assessment of an action at short, mid and long term

A **scenario** is defined as being the *series of views* corresponding to the program of actions. At each intermediaite step of the plan, a state of the forest will be generated and a view will be associated with it. It will be with the aid of these views that the decision-maker will make his choice according to the program of actions to be created.

## SIS as a decision support tool in forestry

### 4.1 Data structure

In order to have the existing SIS evolving towards true Spatial Decision Support Systems, the data, which describe certain components of the decision-making process, must be included in the data structure of the SIS. The resulting decision-oriented SIS may help to preserve landscape quality where wood harvesting is foreseen.

The new elements are called **decision-making primitives** in the same sense as with the geometric primitives (points, lines and polygons) in SIS. These decision-making primitives correspond to the components of the decision-making process described above. Figure 5 shows the general structure of the suggested data model. The shaded boxes correspond to the unforeseeable external factors mentionned above in section 3.2.2. We will neglect these factors since we use a deterministic approach.

The *decision-making primitives* must be integrated as a sub-system into the data structure of the existing SIS. The inventory data are the elements that will connect the new sub-system and the rest of the SIS. However, it is important to underline that the decision-making sub-system will not act directly on the operational inventory data. For reasons of information integrity, the decision-oriented SIS will use *a copy of inventory data*. Furthermore, there should be a link between the two copies since an inventory object may be altered either by the update of the operational inventory or a planned program of actions.

#### 4.1.1 Descriptive and spatial data module

In addition to the copy of the operational data, the *descriptive and spatial* module will contain the inventory data that will result from the planning process. Thus, every state of territory as described above will be memorized here. Also, to enable the decision-maker to manage simultaneously several "future states", the territory objects will be linked independantly to many distinct forest states.

#### 4.1.2 Action-Related data module

This part of the sub-system contains the description of any *modifier* of the forest objects, which are the internal and the external factors of the planning process mentioned in section 3.2.2. The internal factors are organized according to the following structure: *programs - groups of actions - actions* (e.g.: multi-annual plans - annual operation plans - individual interventions). Here, every component is independent in order to allow the decision-maker the entire freedom to create any axis of progression. Thus, an action could be owned by one or many groups of actions; a group of actions could be owned by one or many programs of actions.

#### 4.1.3 State progression

The *descriptive and spatial data* module and the *action-related data* module are connected together by the fact that the components of the second one are acting on the components of the first one. The different links between the two modules are on two levels, for example:

- a *forest object* may be modified by an individual intervention and /or the expected natural growth;
- a *forest state* may be modified by an annual operation plan and/or a multi-annual plan.

#### *4.1.4 Decision-related data module*

This module comprises the elements that will enable the decision-maker to assess the impact of any factor of change and to make decisions. The data that will be stored in this module are *views*, which represent the states of the forest territory at short, mid and long term, and *assessment criteria* that will be applied on the views in order to select the best multi-annual plan.

### 4.2 Planning process

The information that will be contained in the data model described above will be used to create a detailed multi-annual operation plan. In order to do so, the foresters will undertake a planning process which is comprised of three main activities (cf figure 6):

- the *simulation* of a suggested annual operation plan;
- the *extrapolation* over a period of time of the forest states before and after the simulated plan;
- the *assessment* of the views corresponding to these forest states.

Beside the technological aspects and the data manipulation related to this planning process, it can be seen that this approach has an effect on the methods of forest operation planning because these are currently based on a one-year period.

In regards to the integrated management objectives mentioned at the beginning of this paper and the spatio-temporal aspects of the impact of wood harvesting on the forest landscape (Richer, Chevallier 1991), it can be seen that the planning period should be extended.

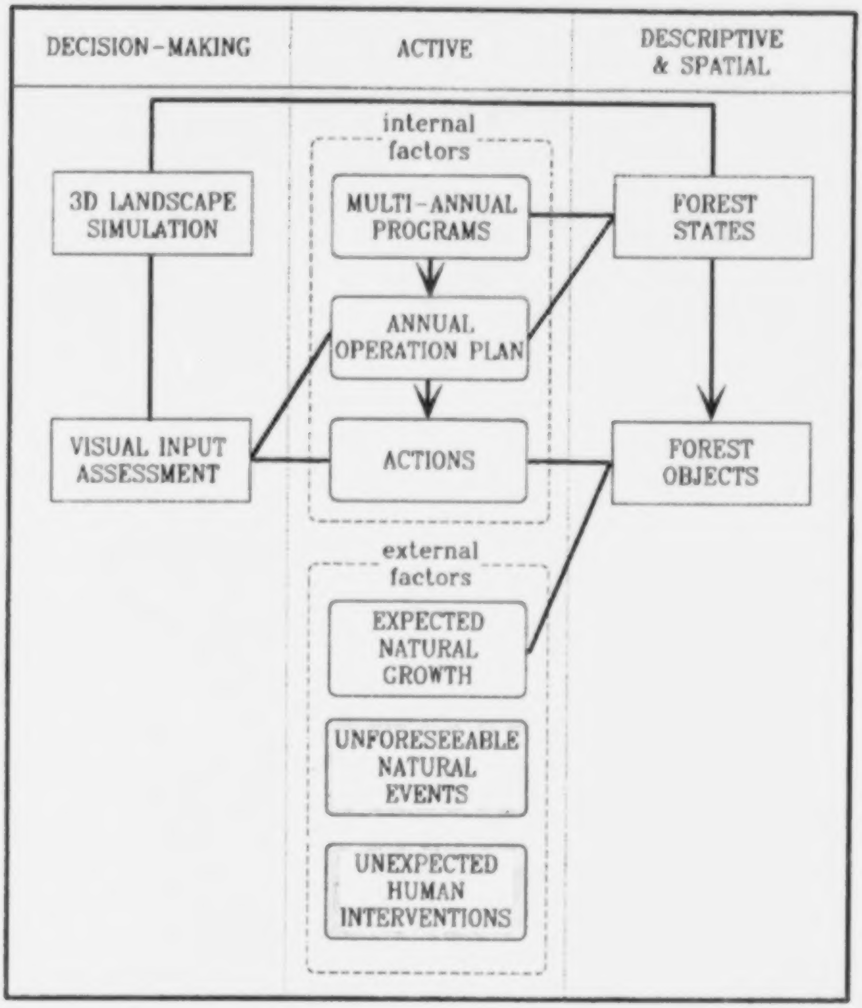

DECISION-MAKING
ACTIVE
DESCRIPTIVE & SPATIAL
internal factors
3D LANDSCAPE SIMULATION
MULTI-ANNUAL PROGRAMS
FOREST STATES
ANNUAL OPERATION PLAN
VISUAL INPUT ASSESSMENT
ACTIONS
FOREST OBJECTS
external factors
EXPECTED NATURAL GROWTH
UNFORESEEABLE NATURAL EVENTS
UNEXPECTED HUMAN INTERVENTIONS

Figure 5 – Data model supporting decision–making in wood harvesting management in regards to landscape quality integrity

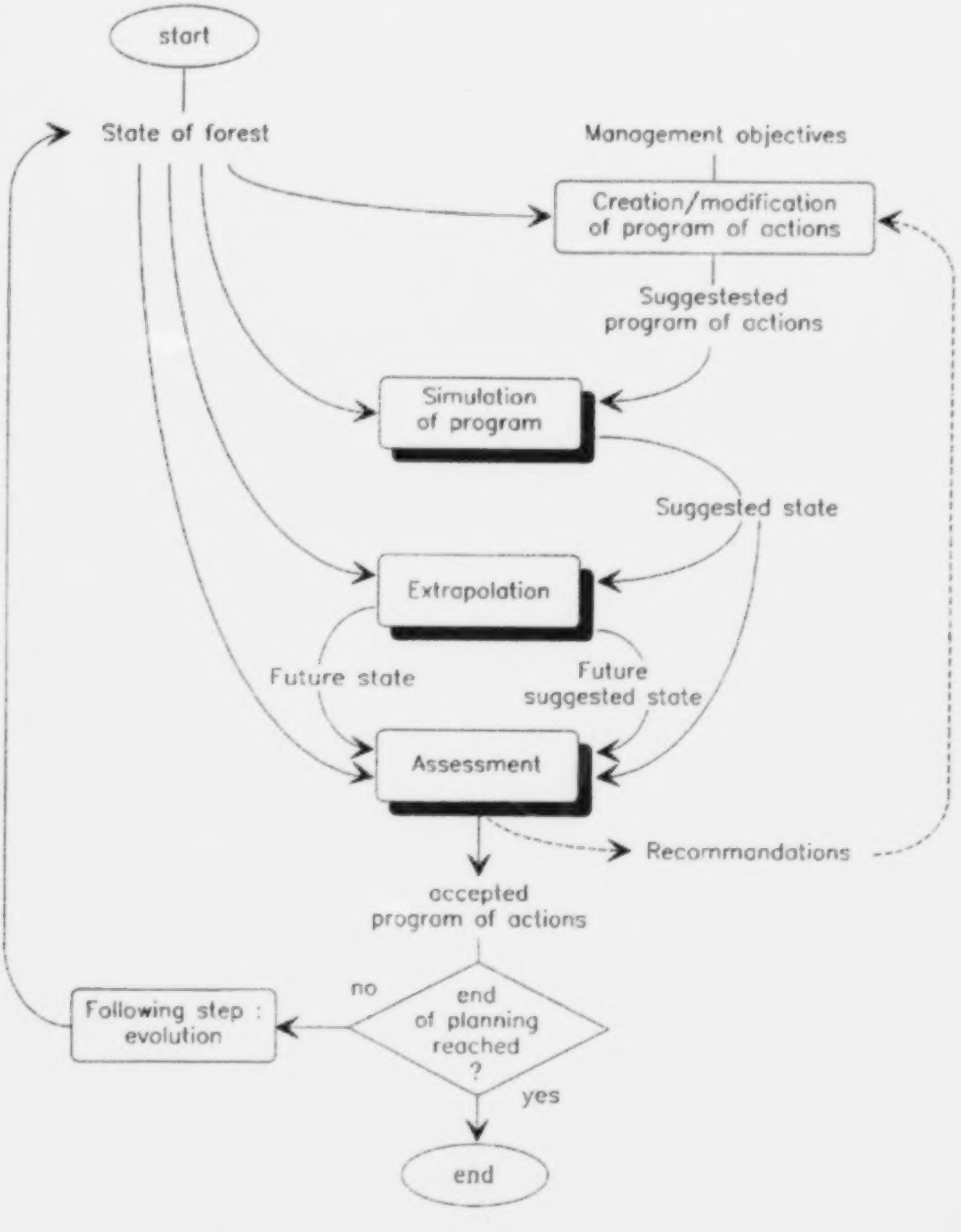

start
State of forest
Management objectives
Creation/modification of program of actions
Suggestested program of actions
Simulation of program
Suggested state
Extrapolation
Future state
Future suggested state
Assessment
Recommandations
accepted program of actions
end of planning reached ?
no
yes
Following step : evolution
end

Figure 6 – Decision-making process

## Conclusion

The evolution of SIS - called *inventory-oriented* for the need of the demonstration - towards true Spatial Decision Support System will require several modifications as much in the software and data as in the work method of the people within the SIS environment. Beyond the simple technonogical progress, it seems evident that this evolution is **necessary** and that it alone will justify the enormous investments put in the collection of inventory data.

In this paper, we have made few steps on the axis of research that aims to enable the decision-maker to benefit more from the use of all technological means. In this perspective, it is likely that future research will take more into consideration human factors and cognitive sciences (Turk 1990) (Chevallier 1992).

## References

CHEVALLIER, J.-J., 1992. *S'informer ou décider: des SIRS aux SADRS in* Proceedings: Canadian Conference on GIS, March 23-26. C.I.S.M. & E.M.R. Ottawa, Canada.

GAGNON, P., 1992. *La gestion du temps dans les GIS: certains concepts fondamentaux in* Proceedings: Canadian Conference on GIS, March 23-26. C.I.S.M. & E.M.R. Ottawa, Canada.

RICHER, O., J.-J. CHEVALLIER, 1992. *SIRS pour la récolte forestière: outil d'inventaire ou moyen de planification? in* Proceedings: Canadian Conference on GIS, March 23-26. C.I.S.M. & E.M.R. Ottawa, Canada.

BILODEAU, J.-M., 1991. *Analyse de la distribution d'une population pour la planification d'échantillonnage appliquée à l'inventaire forestier.* Mémoire de maîtrise, Université Laval, Département des sciences géodésiques et de télédétection. 142 p.

CHAMPOUX, P., 1991. *Étude sur les fonctions d'analyse spatiale à utiliser dans un SIRS appliqué à l'exploitation minière.* Mémoire de maîtrise, Université Laval, Département des sciences géodésiques et de télédétection. 127 p.

CHEVALLIER, J.-J., Y. BÉDARD, 1991. *Échelles de mesure spatiale et processus de modélisation appliqués aux SIRS in* Proceedings: Canadian Conference on GIS, March 19-20. C.I.S.M. & E.M.R. Ottawa, Canada.

LOWELL, K., 1991. *Utilizing discriminant function analysis with a GIS to model ecological succession spatially.* International Journal of GIS. 5: 175-191.

RICHER, O. et J.-J. CHEVALLIER, 1991. *Utilisation d'un SIRS pour la planification des coupes forestières en fonction de la préservation de la qualité du paysage in* Proceedings: Canadian Conference on GIS, March 19-20. C.I.S.M. & E.M.R. Ottawa, Canada.

TURK, A. G., 1990. *Towards an understanding on human-computer interaction aspects of geographic information systems.* Revue Cartography, 19 (1, juin): 31 - 60.

LOWELL, K., 1989. Spatial and temporal GIS modelling using probalistic functions **in** Proceedings: Canadian Conference on GIS, March 19-20. C.I.S.M. & E.M.R. Ottawa, Canada. pp. 1274-1287.

# GIS for visual impact assessment: a case study from the Three Sisters Resort development, Alberta

**David Buckley**

Tel Tech, GIS Division
1430 Florida Avenue, Suite 210
Longmont, CO, USA 80501
Tel: 303-651-2426

## Abstract

*Abstract not provided by authors in time for publication.*

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# B3 Towards the year 2000: FORMAN

**GIS and forest management — towards the next century**

Glen Jordan, University of NewBrunswick, Fredericton, NB, Canada

**GIS and forest management planning — improving operational reality in spatial modelling**

Boyd Pittman, University of New Brunswick, Fredericton, NB, Canada

**GIS and forest mangement planning — integrating strategic and operations considertions**

David Carleton et al., University of New Brunswick, Fredericton, NB, Canada

**Up the GIS learning curve — the New Brunswick forest mangement experience**

Bob Dick, Department of Natural Resources and Energy, Fredericton, NB, Canada

G I S
92

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# GIS and forest management — towards the next century

**Glen Jordan, RPF**

Associate Professor
Faculty of forestry
University of New Brunswick
Bag Service Number 44555
Fredericton, NB E3B 6C2

### Abstract

*This paper assesses the present sophistication of GIS activity in the Canadian forestry sector and suggests a framework for research and development of the technology's full potential in forest management. It begins with an explanation of the rationale behind the strong forestry — GIS association in Canada, followed by an elaboration on GIS and forest management. The paper notes that the location-dependent nature of forestry decision-making, explains widespread GIS activity and interest. It is revealed, however, that current activity is mostly limited to inventory (store and query) applications with the full potential of GIS mostly untapped in the other cornerstones of forest management: management design, on-the-ground implementation of forest management interventions, and assessment of forest response to interventions. Improving the effectiveness and efficiency of forest management decision-making through research and development of GIS-based data and data processing links among these four cornerstones is presented as the challenge of the 90s. The paper concludes with a description of a research and development initiative at the University of New Brunswick that is so directed.*

*Keywords: GIS, forest management, research*

## Introduction

Geographic information system technology is widely used in the Canadian forestry establishment, and for good reason. Appearing on the scene commercially more or less ten years ago, GIS technology experienced rapid adoption by first government agencies and then industrial concerns across the country (see Figure 1). At present, nine of ten provinces have province-wide GIS-based forest inventories in one stage of development or another and industrial initiatives too are numerous. While the forestry sector is no longer the sole champion of GIS technology application in Canada, forestry can be credited with proving the technology and having hastened its now widespread adoption. The Canadian forestry sector does, however, retain the distinction of having the longest association with GIS. Beginning with a flurry of benchmark and acquisition studies in the early 1980s, GIS application has evolved to its present state. This strong and long forestry GIS association isn't surprising. Planning and decision-making in most forestry operations in Canada take place in a very large geographic domain. It's not uncommon for forestry companies to own or hold licence to forests of one million hectares or more. Map-based data are prerequisite to efficient mining, not to mention management, of such large forests.

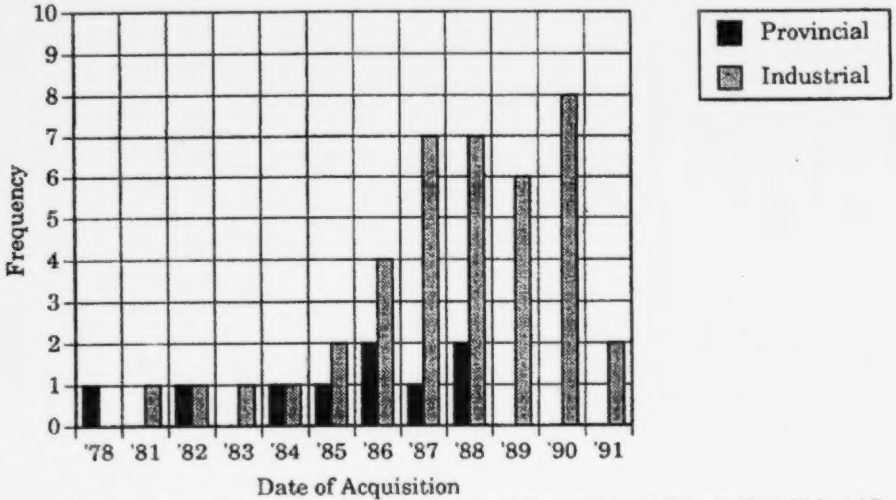


**Figure 1:** *Chronological distribution of provincial and industrial GIS acquisitions in Canada (from Dick and Jordan, 1990).*

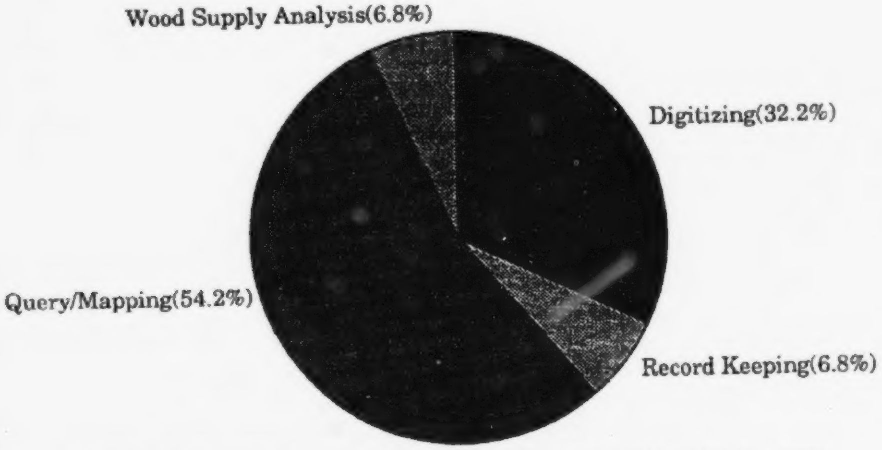


**Figure 2:** *Distribution of forestry GIS activities in Canada (from Dick and Jordan, 1990).*

This paper assesses the present sophistication of GIS activity in the Canadian forestry sector and suggests a framework for research and development of the technology's full potential in forest management. It begins, however, with an elaboration on the rationale behind the strong forestry — GIS association, followed by a discussion of GIS and forest management.

## Whither GIS in Canadian forestry?

Forestry requires data that describe the forest resource in terms of both conditions and their geographic distribution. These spatial data typically consist of measurements that describe the physical condition of a forest's component stands and covertype maps that define their geographic location.These two components of spatial data typically form the core of forest inventories, and their processing the basis of forestry planning and operations. It isn't surprising, therefore, in this time of rapidly advancing technology development, that forestry has generated a great demand for computer-based, data-handling systems capable of accommodating spatial data. GIS technology obviously has potential in forestry since it is unique among information systems in providing the capability to store and process spatial data. It's no coincidence that the province with the longest and most successful application of GIS (Dick, 1990) achieved the highest forest management rating in a national survey conducted by MacLean's Newsmagazine (Anon., 1991).

GIS activity in Canadian forestry, however, is relatively unsophisticated in light of the technology's potential. The creation, maintenance and production of covertype maps and conduct of simple map-based queries on forest inventory have, to date, characterised GIS application. Results of a survey of GIS use in Canadian forestry by Dick and Jordan (1990) confirm as much. Across the almost fifty government agencies and industrial firms identified in the survey as having a forestry mandate and GIS capability, the majority of their activities focused on digitizing, mapping and simple map-based queries, and harvesting and silvicultural record keeping (see Figure 2). Very little GIS activity was reported in more complex levels of application in forest management, such as wood supply analysis. The issue now is not the current state of GIS in Canadian forestry but, rather, capture of the technoloy's full potential in the future. Understanding the forest management process will prove crucial in this regard.

## Putting management into forestry

Canadian forestry has only recently begun to show signs of making the transition from strictly harvest management to forest management. While harvest management focuses on short-term economic control of timber extraction, forest management takes a longer-term outlook and focuses on control of forest development. The two are not mutually exclusive. Harvesting, in combination with other interventions, is a powerful tool of forest management (Baskerville, 1986). A problem arises, however, when harvest and other intervention management operate separate from, or in the absence of, a long-term forest perspective. Such activity is not forest management and unacceptable in today's world.

Forest management is a process aimed at long-term control of forest inventory. It attempts to accomplish control by designing, implementing and monitoring a set of interventions in which forest areas are harvested, products are distributed, silviculture is applied, and protection against insects, fire and disease is provided (Baskerville, 1986). These activities are timed and located so that their cumulative effect controls forest development and provides an inventory with the desired mix of benefits or values (eg. timber, habitat, recreation opportunity) across the whole forest over time. The forest management challenge is to: (i) find one schedule, i.e. design, that appears likely to produce the desired forest development pattern and inventory; (ii) implement that schedule year-by-year; and, (iii) monitor forest inventory periodically to look for and remedy divergence between expected and actual outcomes (Jordan and Erdle, 1990).

The first challenge requires projection, i.e. forecast, of present forest inventory into the distant future in response to timed and located interventions. With this it is possible to heuristically identify one schedule, from among the many possibilities, that appears best. The second challenge implementing the intervention schedule on the ground requires that yearly harvesting, harvest distribution, silviculture and protection activities be assigned appropriate equipment and personnel, seasonally linked and efficiently deployed. The third challenge monitoring forest inventory requires the comparison of actual forest response against that originally forecast for the schedule selected.

Collectively, ***inventory, design, implementation and monitoring*** form the cornerstones of the forest management process (see Figure 3). An up-to-date description of forest *inventory* is an essential starting point in the management process. Forest modelling is a key ingredient in inventory projection and forest management *design*; cost-effective linking (over time and landscape) of interventions is a key concern in forest management *implementation*; and an accurate, up-to-date log of intervention and growth responses is essential in *monitoring* cumulative forest performance (Jordan and Erdle, 1990).

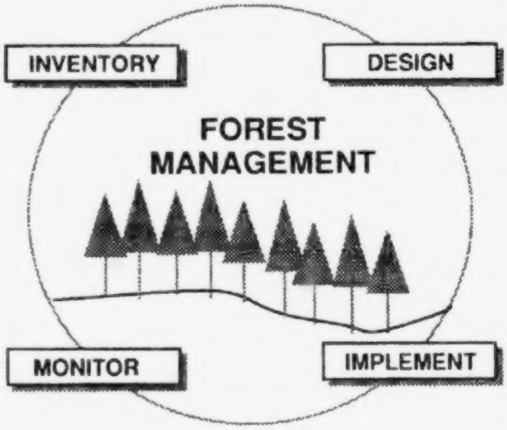


**Figure 3:**
*The four cornerstones of the forest management process.*

## Putting GIS to work in forest management

A major challenge for Canadian forestry in the 90s will be realizing fully the enabling effect of GIS technology in the forest management process. As already noted, GIS activity in Canadian forestry is mostly limited to inventory, i.e. store and query, applications. The greatest potential of GIS lies untapped, almost totally, in the other cornerstones of forest management: design, implementation and monitoring. Realizing this potential will involve researching and developing a new level of

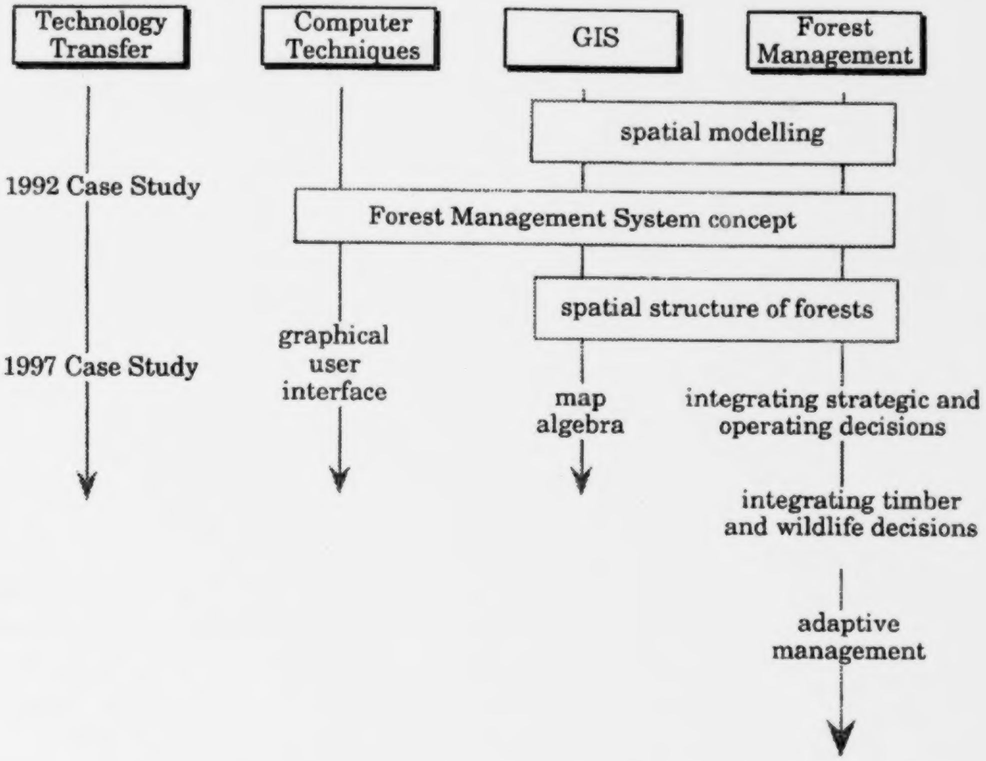

Technology Transfer
Computer Techniques
GIS
Forest Management
spatial modelling
1992 Case Study
Forest Management System concept
spatial structure of forests
1997 Case Study
graphical user interface
map algebra
integrating strategic and operating decisions
integrating timber and wildlife decisions
adaptive management

**Figure 4:** *The FORMAN 2000 Project has four streams of broad areas of activity.*

computer-based technology, forest management systems (FMS). This will be achieved in much the same way as the researchers and developers of GIS have expanded database management system (DBMS) technology into the spatial data domain. Forest management systems will build upon and extend GIS (and other computer technologies).

Forest management system technology will permit the establishment of data and data processing links among the four cornerstones of the forest management process. A GIS will supply the principal, although not the only, source of data and means of data processing. A GIS database will store: (i) present, up-to-date, inventory (including growth and yield) and future inventory; (ii) long-term management, i.e. intervention, schedules; (iii) short-term operations plans; and, (iv) an accumulated log of actual interventions. GIS data processing will include: (i) spatial forest modelling for forecasting and management design; (ii) accessing, editting and revising of management schedules in operations and protection planning; (iii) periodically reconciling actual forest inventory with forecast future forest inventory; and, (iv) continuous inventory updating. This integrated database and data processing approach will link inventory, design, implementation and monitoring, and improve both the efficiency and effectiveness of the forest management process.

## The FORMAN 2000 project

The FORMAN 2000 Project, recently initiated at UNB's Faculty of Forestry, is based upon the FMS concept. With initial three-year funding from Forestry Canada, NSERC and four Atlantic Canada forest industries (Corner Brook Pulp and Paper Ltd, Fraser Inc, J.D. Irving Ltd and Juniper Lumber Ltd), the Project aims to develop a working FMS by the year 2000 and expand the science of forest management as well. The Project will draw strength from: (i) New Brunswick's highly regarded framework of forest management; (ii) a forest management and technology-wise community of foresters in New Brunswick; and, (iii) a strategy of drawing together industry, government and academic stakeholders.

To date the Project has focused on management design and implementation. A GIS database and data processing tools that help bridge the two cornerstones are the principal focus of research and development at present. So far, efforts have lead to the conceptual development and programming of a GIS-based forest model and forest management design process. Unlike aspatial counterparts, a spatial model uses the absolute and relative positions of stand conditions kept in a GIS database to temporally and geographically link harvest scheduling, silviculture scheduling, harvest distribution and protection in management design. The Project's current spatial model employs GIS data to mimic harvest blocking, including greenup, and production of location-specific, map-based harvest and silvicultural schedules (Baskent and Jordan, 1991). Unlike contemporary approaches to development of management plans, a GIS-based approach produces a long-term schedule of interventions, i.e. harvest and silviculture, in digital form, readily available for display, operations planning, monitoring and revision.

The Project's 10-year research and development agenda involves four interdependent activity streams: technology transfer, computer techniques, GIS and forest management (see Figure 4). A full-time research assistant ensures that the Project's development agenda, i.e. transfer of technology to industrial partners, is executed successfully. In this stream, the Project faces two important milestones: New Brunswick's 25-year forest management plan revisions scheduled for 1992 and 1997. Both will provide testing grounds for Project developments. The principal interest area in the GIS stream is development and application of raster processing techniques, i.e. map algebra. The computer techniques and forest management streams will involve participation of both graduate students and industrial partners. Graduate student research ranges from conceptual, PhD-level undertakings to technically-oriented masters-level initiatives and covers broad areas of investigation such as graphical user interfaces, forest management system conceptual development, operational considerations in management design, spatial forest modelling for timber and wildlife, adaptive management and spatial structure in forests.

## Summary

Although GIS technology is well established in Canadian forestry, its application lacks sophistication specifically, in forest management. A major challenge for the 90s will be addressing not only the present scarcity of forest management in Canada but an apparent lack of GIS know-how beyond simple store and query applications. The latter necessitates a significant research and development initiative within an appropriate framework. Modelled on the forest management process, the framework suggested is directed toward a new level of computer technology forest management systems that will permit the establishment of data and data processing links among the four cornerstones of the forest management process: inventory, design, implementation and monitoring. The FORMAN 2000 Project at UNB is an academic, industry and government initiative based on the framework.

## Literature cited

Anon. 1991. Rating the forests. *Maclean's Newsmagazine,* December 16, pp 44-45.

Baskent, E.Z. and G.A. Jordan. 1991. Spatial forest simulation modelling. *Forestry Chronicle.* In press.

Baskerville, G.L. 1986. Understanding forest management. *Forestry Chronicle* 62:339-347.

Carleton, D. 1992. GIS and Forest Management Planning — Integrating Strategic and Operations Considerations. In Proceedings of GIS '92 Symposium.

Dick, R.C. 1990. GIS: The right stuff. *In* Forest Management and Geographic Information Systems — Volume I, R.C. Dick and G.A. Jordan ed., pp. 131-137.

Dick, R.C. and G.A. Jordan. 1990. GIS activity in Canadian forestry. *In* Forest Management and Geographic Information Systems — Volume I, R.C. Dick and G.A. Jordan ed., pp. 144-176.

Jordan, G.A., and Erdle, T.A. 1989. Forest management and GIS: What have we learned in New Brunswick? The Canadian Institute of Surveying and Mapping Journal, 43(3): 287-295.

Pittman, B.1992. GIS and forest management planning — improving operational reality in spatial modelling. *In* Proceedings of GIS '92

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# GIS and forest management planning — improving operational reality in spatial modelling

**Boyd Pittman**

MScF Student

**Rick Wightman**

Research Assistant
Faculty of Forestry
University of New Brunswick
Fredericton, N.B.

## Abstract

*This paper reports on proposed research that will extend spatial forest modelling. It begins with a brief review of the GISFORMAN spatial model. This model, developed at UNB, has established a framework for spatial modelling and the inclusion of timber cost in forecasting wood supply and harvest schedules. Using stand locations and relative positions stored in a GIS database, the model is able to include a component of timber cost influenced by geographic distribution of forest conditions in the determination of wood supply. However, the paper notes that a significant additional component of timber cost attributable to geographic distribution of forest conditions, still remains unaccounted for. Specifically, GISFORMAN computes an operating cost* ($\$/m^3$) *for timber by mimicking actual harvest operations and concentrating harvest into economic harvest blocks by clustering stands and computing cost by weighing harvest block road construction costs against volume content. However, lacking any means to control the temporal concentration of cut blocks geographically, the model ignores an important economic reality in harvest scheduling. The paper offers an approach for incorporating this additional component of timber cost, i.e. the cost of linking harvest blocks to a main haul road. In a fashion similar to grouping stands into harvest blocks to concentrate volume and control costs, it is proposed that harvest blocks be grouped into harvestsheds to control access costs.*

*Keywords: access costs, forest management, forest modelling, GIS, harvest scheduling, timber economics, spatial constraints*

## Introduction

Forest management seeks to ensure a sustainable flow of wood resources from a forest. The premise of forest management is that actions taken, or not taken, today in a forest determine its capability to supply benefits in the future. In Canada, wood supply simulation modelling techniques have played an important role in evaluating intervention strategies and establishing forest management plans. The models help predict the influence of planned management actions on future wood supply.

Wood supply is a function both biological and spatial availability. Biological supply is a function of timber maturity, i.e. operability, while spatial supply is a function of access cost limitations. Within any time frame or period, operable timber supplies are available only if the cost of accessing them does not exceed allowable cost limits. If a wood supply model provides no mechanism for gauging the effect of spatial arrangement of harvest blocks on access costs, it is possible that while the forecasted level of harvest may be biologically sustainable in the future, it may not be economically sustainable. There is the inherent danger of scheduling all the inexpensive wood today, thus leaving only expensive wood for the future. It is imperative that spatial constraints on harvesting operations be mimicked in wood supply forecasting.

A spatial wood supply model called GISFORMAN has been developed at the University of New Brunswick (Baskent, 1990). GISFORMAN links FORMAN (Wang et. al., 1987), an aspatial wood supply model commonly used in Atlantic Canada, to a GIS database (ARC/INFO). Development of GISFORMAN focused on adding location-specific resolution to FORMAN. FORMAN circumvents its lack of spatial parameters by aggregating similar stands into stand-type classes. The stand-type is the base unit within the model, in which all activities are scheduled. The output from the model simply consists of a listing of the area (ha) to be harvested by stand-type. In operational forestry, however, harvest blocks are the base unit for management actions for economic reasons. A block is defined as a group of stands, of similar species content and stage of development, that are either adjacent to each other or dispersed within an area small enough so that the stands can be treated as a unit. Much work is needed to transform tabular output produced by FORMAN into blocks that can be drawn on operational maps. Furthermore, there is no guarantee that it will be possible to group stands of the stand-types scheduled by FORMAN into operationally feasible harvest blocks. GISFORMAN increases the resolution of FORMAN by utilizing location specific harvest blocks instead of stand-types. GISFORMAN uses the topological data stored in a GIS database to group adjacent stands into harvest blocks. These blocks are then scheduled for harvest using one or more harvest rules. Harvest rules define the criteria used in queueing blocks for harvest.

The location-specific resolution of GISFORMAN offers several advantages. The model produces maps displaying harvest blocks to be harvested within a period, saving a planner considerable time. However, because the model defines harvest blocks using stand boundaries, some squaring or generalization of the "pseudo harvest blocks" outlined by the model may be necessary in order to construct operationally feasible blocks. Another strength of GISFORMAN is its ability to incorporate the operating cost of harvesting a block into wood supply modelling. Within a block, the model computes an operating cost ($\$/m^3$) by weighting the cost of building operational roads against volume content of the block. If the operating cost exceeds a set limit, the block will not be considered for harvest. The ability to adjust the allowable operating cost provides a mechanism for evaluating the impact of different operating road cost limits on sustainable wood supply. Further, GISFORMAN mimics opera-tional forestry by spatially constraining harvest activity to blocks. In practice, when operational blocks are formed, some stands are harvested before they have reached their peak or maximum volume. Blocking thus has the negative side effect of lowering sustainable wood supply. GISFORMAN, because of its ability to mimic harvest blocking, can account for the effects of blocking on wood supply.

While GISFORMAN has the ability to group stands into harvest blocks, it lacks the ability to control the temporal, spatial pattern of blocks across a forest. GISFORMAN presently assumes unhindered access for all harvest blocks and, as a result, produces at any time harvest blocks scattered throughout the forest (Figure 1). The whole forest may not be physically accessible, however, at any given time due to lack of a complete road network. Further, even with full access, it may not be economically desirable to scatter harvesting operations over the whole forest.

This paper will introduce a research project recently initiated at the University of New Brunswick that will address this shortcoming. Conceptual development of an improved spatial model will be presented and its implications discussed.

## The proposed project

The project will address the problem of spatially constraining harvest blocks within areas that are physically and economically accessible by road. GISFORMAN will provide the core for the development of the proposed model. Using the idea of a "harvestshed", originally suggested by Berry and Sailor (1981), harvest blocks will be grouped into zones having a central forest access road for their nucleus. The model will provide a mechanism to allow the manipulation of criteria defining physical and economical accessibility, thereby permitting the evaluation of these access constraints on sustainable wood supply. Physical accessibility is determined by the presence of a permanent forest access road. By specifying the amount of permanent road to be constructed within a period, the planner can control the level of physical access by specifying the number of harvestsheds available. Economic accessibility is dependent upon the access cost (\$/m3) within a harvestshed. Access cost is defined as the cost of building temporary roads that link harvest blocks to a permanent road. The amount of temporary road needed is a function of the number and size of harvest blocks within a harvestshed and the volume of wood within the harvest blocks. By varying the amount of permanent access road to be built within a period, and the maximum allowable access cost limit ($\$/m^3$),

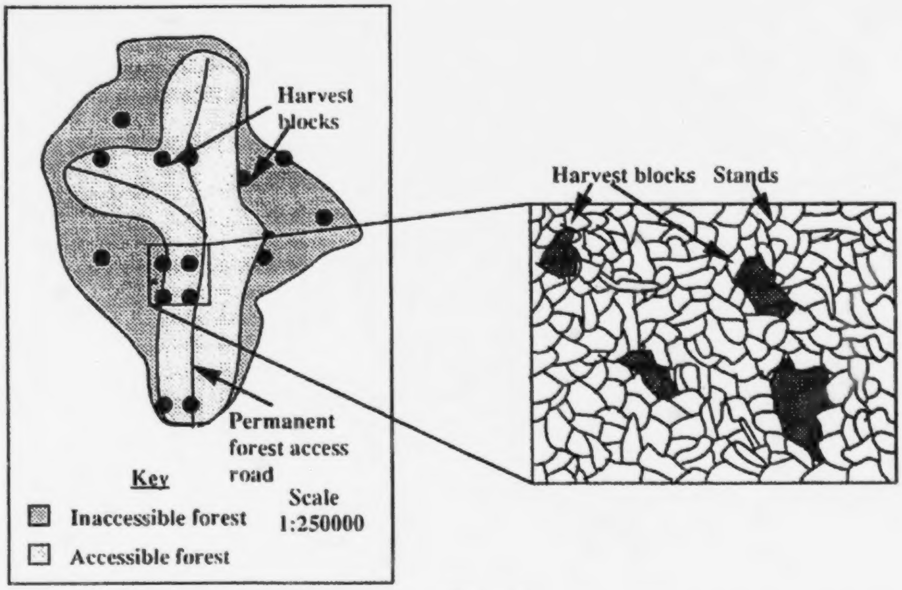

Harvest blocks
Harvest blocks
Stands
Permanent forest access road
Key
Inaccessible forest
Accessible forest
Scale 1:250000

**Figure 1:**
*GISFORMAN groups stands together forming location specific harvest blocks. The model lacks the ability to control the spatial pattern of harvest blocks across the forest resulting in harvest blocks being scheduled independent of blocks having access.*

it is possible to measure the impact, in terms of sustainable wood supply, of different access strategies.

The physical extent of a harvestshed will be determined by two factors: (i) the length of the permanent access road that forms the central axis of the harvestshed and (ii) the average length of temporary road access that radiates out from the permanent access road to the blocks within the harvestshed. The inclusion of a block in a harvestshed will be determined by the distance from the center of the block's center-most stand to the permanent road. If this distance is less than the specified average length of temporary road, the block will be included.

The two assumptions that form the core of this research are: (i) full access to all blocks is not available, and (ii) with time the road infrastructure will increase. However, the scope of the model is not to position roads on the landscape, but to constrain blocks to areas that have roads in place or areas in which road building is proposed. Therefore, the model requires a map containing both existing roads and those roads that will exist when the road infrastructure is completed. For each period, the model will create a harvestshed around each proposed road segment attached to the existing road network. Road segments flagged by the planer for construction (pre-scheduled) in that period will be opened first, regardless of their cost. After pre-scheduled road segments have been opened, additional segments will chosen based on planner specified ranking criteria.

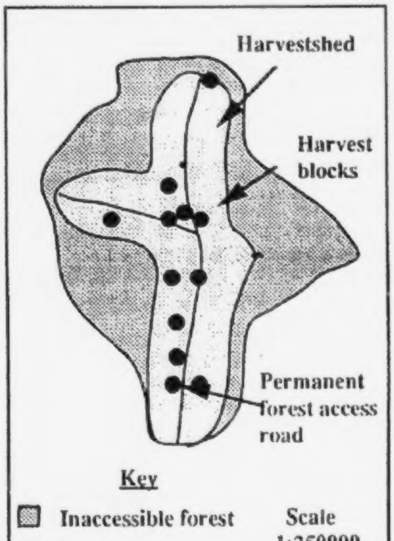

Harvestshed
Harvest blocks
Permanent forest access road
Key
Inaccessible forest
Accessible forest
Scale 1:250000

**Figure 2:** *The proposed model will permit the spatial pattern of harvest blocks to be controlled within a forest. Harvest blocks will be spatially constrained to areas (harvestsheds) accessible by road.*

The proposed road segments that are associated with the highest ranking harvestsheds will be scheduled for construction. Opening of roads based on their associated harvestshed ranking will continue until the planner's specified limit on road construction for a period is met.

After the road segments to be constructed within a period have been selected and added to the permanent road network, the model will then rank all harvestsheds that are accessible by permanent road network based on criteria (harvest rules) specified by the planner. This ranking will include all the new harvestsheds as well as the accessible harvestsheds that existed previously. As with GISFORMAN, there will be harvest rules that permit the planner to control the criteria for queuing harvest blocks; however, now the proposed new harvest rules will rank harvest blocks within a harvestshed and not the whole forest. By ranking blocks within the harvestshed, the harvestshed can be also ranked based on the collective harvest block values they contain.

While GISFORMAN offers three spatial harvesting rules, this project will add three more. The first rule will minimize unharvested volume loss within a harvestshed. Blocks whose component stands are near the end of their yield curve, and are expected to experience volume losses before the next period, will be placed at the front of the harvest queue. The model will select blocks from this queue until all blocks have been scheduled or the mortality ($m^3$/ha) within a block falls below a planner defined level. The second rule will maximize volume harvested within a harvestshed. Blocks that have the greatest concentration of volume will be placed at the front of the harvesting queue. As with the first rule, the model will select blocks from the queue until volume concentration ($m^3$/ha) within a block falls below a planner defined level. The third harvest rule will minimize access cost within a harvestshed. Access cost will be determined by the cost of constructing temporary roads. The amount of temporary road (km) needed within a harvestshed will be controlled by the size and number of blocks that are to be harvested within the harvestshed. The planner must specify the average amount (ha) of a harvest block area that can be accessed by a kilometre of temporary road. This value will be based on the planner's knowledge of the terrain and the road building density specifications used within the forest. The total temporary road building cost for the harvestshed will be the sum of the cost of building temporary roads to each block scheduled. This total road building cost will be weighed against the volume removed from the harvestshed.

Other factors associated with the cost of accessing a block, such as construction and maintenance costs for permanent access roads, will not be attributed to a single harvestshed. Since the amount of permanent road to be constructed in a period will be determined by the planner, it will be viewed as an investment in the road network infrastructure and charged against all wood harvested in that period. The maintenance costs of a permanent road also cannot be solely attributed to the volume of wood removed from the harvestshed because all permanent roads will receive maintenance while only a select few will support harvesting in any one period. Furthermore, a segment of permanent road will serve other uses such as fire protection within a period, whether or not wood is harvested from the harvesteshed it defines. Therefore, the totoal cost of maintenance will be charged against all wood harvested in a period without regard to harvestsheds.

## An example of how the model will be used

Upon completion of the model, it will be used to investigate forest dynamics and impact on sustainable wood supply of different access strategies. For this purpose a 1440 km2 study area situated in the Little Grand Lake area of Western Newfoundland will be employed. The area is composed primarily of the land owned or licensed to Corner Brook Pulp and Paper Limited. To date, wood supply analysis with FORMAN has revealed that maximizing sustainable wood supply depends primarily on minimizing volume loss due to mortality. The proposed model will be used to spatially constrain harvest blocks to harvestsheds that are accessible by a permanent forest access road. Although the sustainable harvest level is likely to fall as a result of spatially constraining harvest blocks, the resulting projection should be more realistic and closer to the "true" sustainable harvest level. Furthermore, given that further development of the present road network will occur over the next 20 to 25 years, the amount and timing of permanent access road construction can be varied on a period by period basis, and the maximum allowable access cost limit can be varied to create different access strategies.

Different access strategies will be developed and the resulting access costs weighed against a desired level of sustainable harvest. For example, one possible strategy could be to open up all harvestsheds within the study area in the first two periods instead of spacing the level of road construction evenly across four to five periods. It is expected that opening all harvestsheds within the first two periods will help maximize the sustainable harvest by improving the ability to capture harvest blocks that will experience high volume loss due to mortality. However, opening up all the potential harvestsheds will increase the access costs within the first couple of periods, and these higher access costs may exceed acceptable limits. By varying the number and timing of road construction projects, it will be possible to determine a strategic road building schedule that will not exceed an allowable access cost limitation (\$/$m^3$) and will generate an acceptable sustainable level of harvest.

## Summary

When determining wood supply, it is not only imperative to have sufficient wood available, but the wood supply must be geographically located such that it is physically and economically accessible. To realistically model wood supply and schedule harvesting operations it is necessary to differentiate between the effective forest (i.e. the forested area that can be physically and economically accessed and utilized at any given time) and the total land base. The FORMAN and GISFORMAN models were designed with the assumption that the whole forest was fully accessible. Therefore, they lack the ability to differentiate between the whole forest and the effective forest. The research project introduced in this paper will attempt to incorporate a mechanism into GISFORMAN that will spatially constrain harvest blocks to areas that are physically and economically accessible by road. Harvest scheduling with this ability should permit a more realistic forecast of future forest development and evaluation of different access strategies on sustainable wood supply.

## Literature cited

Baskent, E.Z. 1990. Spatial Wood Supply Modelling: Concept and Practice. M.Sc.F. Thesis, University. of New Brunswick, Fredericton, N.B. 85 pp

Berry, J.K. and Sailor, J.K. 1981. A spatial analysis of timber supply. In In-Place Resource Inventories: Principles and practices, University of Maine, August 9-14, 1981, pp 823-833.

Wang, E.C., Erdle T.A., and Roussel T. 1987. FORMAN wood supply model user manual. Version 2.0, NBFRAC, Fredericton, N. B., pp

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# GIS and forest management planning — Integrating strategic and operations considerations

**David Carleton**

MScF Student

**Glen Jordan, RPF**

Associate professor
Faculty of forestry
University of New Brunswick
Fredericton, N.B.

## Abstract

*In Canada, a significant portion of the forest management planning process involves the use of computer models to forecast future forest inventory, in response to alternative harvest levels and strategies. However, none of the models take into account the many on-the-ground constraints that an experienced forester must consider while creating an actual harvest block schedule. As a result, harvest schedules produced by the models need considerable editing and the resultant negative impact on sustainable harvest level assessed. This paper describes an enhancement to the planning process, using a New Brunswick case study, whereby a link has been established to operational reality in the management design process by creating an operations feedback loop in harvest scheduling. The feedback loop could be implemented using any contemporary model-based planning process that uses digital forest inventory. The GISFORMAN model, however, is used in this investigation since, unlike most current forest models, it is spatial and provides geographic control of the distribution of a harvest across inventory types while building and mapping a harvest block schedule. The paper describes some of the basics of the GISFORMAN model and, more importantly, how the model and planning process have been modified in an attempt to incorporate an operations feed-back mechanism.*

*Keywords: forest management, GIS, operations, harvest scheduling, spatial model*

## Introduction

In New Brunswick, the Crown Lands and Forests Act of 1980 places the onus for forest management on the provincial government and ten Crown Forest Licensees (Ker, 1987). The forest management plan prepared by each Licensee to fulfill part of their obligation under the act, is a 25 year plan that is updated every 5 years. In the plan, the Licensee must address; (i) species, volume, and quality of standing timber, including natural and artificial regeneration, (ii) stand growth rates and silvicultural methods to be employed (this includes the type, timing and level of silvicultural activity), and (iii) locations and quantities of product to be harvested by operating season.

Figure 1 is a simple illustration of the planning process used by Crown Forest Licensees in New Brunswick. The process is not unique to New Brunswick and could be applied by any organisation to manage any forest.

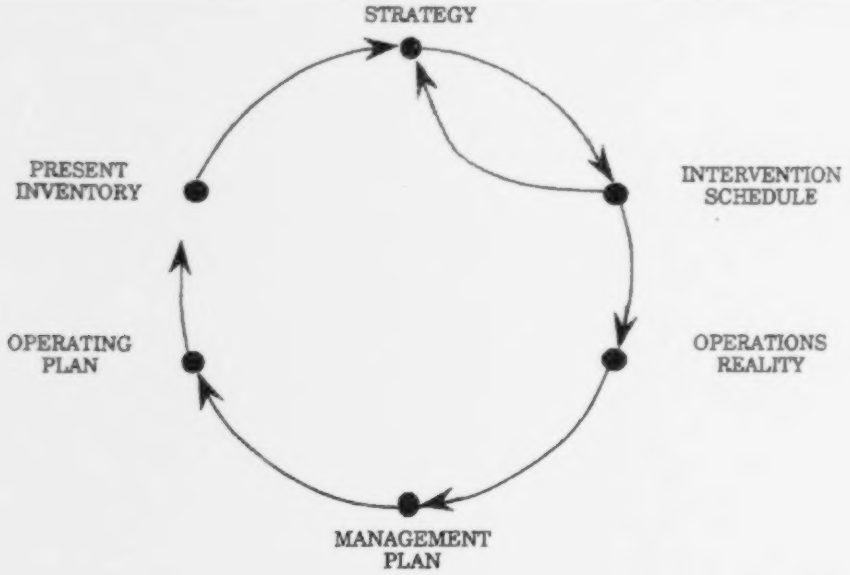


**Figure 1:**
*A typical forest management process illustrating the absence of a feedback loop between strategic and operations considerations.*

It exhibits, however, the typical lack of a strong link and absence of a feedback loop between strategic and operations considerations. The New Brunswick Crown Forest Licensees used the process to develop their 1987 management plans. As part of the process, the FORMAN (Wang et al, 1987) wood supply model assisted in developing harvest schedules. Change in the process for 1992 will occur since Licensees will have access to improved versions of the FORMAN model: FORMAN+1 (Vanguard, 1991) and GISFORMAN (Baskent, 1990).

A shortcoming of FORMAN, FORMAN+1, and other aspatial wood supply models, is that they lack the capability to spatially allocate, i.e. map, the schedules produced. Aspatial models use forest aggregates or classes (hectares of common development type), as opposed to stand level information, to first, forecast development of the forest, and second, determine a sustainable harvest level. The aspatial harvest schedules that Licensees created in 1987 using FORMAN, required "translation" into stand-by-stand schedules. Staff at the Department of Natural Resources and Energy (DNR) completed this task for Licensees by using the Department's GIS to break out stands comprising forest aggregates (Erdle and Frame, 1991). The result was a series of harvest schedule maps showing the 5-year harvest period for each stand targetted for harvest. The scheduled stands were then manually grouped by the Licensees to form operational harvest blocks, primarily on the basis of economics (proximity to existing road networks and mills), but also in an attempt to minimize mortality losses and maximize volume harvested. Unfortunately, accurate harvest volume statistics associated with the manually blocked harvest schedule were not produced. Determination of harvest level was by visual inspection and approximation. That is, an estimate was made of the percentage contribution of each harvest block stand to the block's total harvest volume. This process was required for every block laid out for a license a labourious and inaccurate process. Due to the time involved, investigation of alternative blocking patterns was not possible.

A recent development in GIS-based, or spatial, wood supply modelling has created an opportunity to improve the management planning process in New Brunswick. The GISFORMAN model takes a GIS-based forest inventory, forecasts development of the forest over time, and creates a harvest schedule that can be mapped easily. Furthermore, stand attributes such as maturity class and species composition, are used by the model to form pseudo-blocks (neighbourhoods of similar stands) that can then be used as a guideline for manual harvest blocking. Even though the planning process could be significantly more automated, i.e. shortened, using GISFORMAN, consideration of many operational factors is still lacking and thus the capability to accurately determine a sustainable harvest level remains impaired.

Currently in New Brunswick, the Crown Forest Licensees are going through the process of updating their 25-year management plans. Three Licensees, J.D. Irving Ltd., Fraser Inc., and the Juniper Lumber Co. Ltd., will make use of the GISFORMAN wood supply model to assist in determining harvest and silviculture schedules. The remaining Licensees will use only the aspatial FORMAN+1 model and, therefore, a planning process similar to that used in 1987.

This paper will discuss recent enhancements to and planned application of the GISFORMAN model in 1992 management planning in New Brunswick. The enhancements provide a link between strategic and operations considerations in the planning process by providing the capability to measure the effect of harvest blocking on sustainable harvest level and to modify the harvest block schedule should the result be unacceptable. All three Licensees aforementioned will make use of this additional functionality, but Juniper Lumber Co. Ltd. will be the main test case and it is their situation that will be discussed in detail.

## The GISFORMAN model

The GISFORMAN model is a GIS-based, or spatial, wood supply model developed at the University of New Brunswick. What distinguishes this model from other contemporary wood supply models is: (i) the ability to forecast forest development based on stand level information, and (ii) the ability to incorporate the geographic position of stands into the testing of harvesting strategies and development of harvest block schedules (Baskent, 1990). The current GISFORMAN model makes use of a front-end process (SCAN routine) that identifies neighbouring stands for each stand of a subject forest based on similarities in user-defined thematic attributes. Thus, for a forest comprised of 1000 stands for example, 1000 neighbourhoods would be created. Then, to test harvest levels and strategies, GISFORMAN creates pseudo-blocks by analyzing each neighbourhood in terms of contribution to requested harvest level and compliance with the given strategy. There are two types of output from GISFORMAN. The first is a report detailing the inputs used and the results of the wood supply analysis (periodic levels of harvest, planting, and thinning, area cut, growing stock condition, etc.). The second is a dataset containing the identification and location of stands to be harvested in each period (number of periods and period length are user-defined). GISFORMAN maintains an ARC/INFO dataset in the modelling process and through the use of the ARCPLOT subsystem, maps may be generated showing location of cutting, planting, and thinning by period of application.

A mapped schedule from GISFORMAN provides a forest manager with better information than does the tabular schedule produced by aspatial wood supply models such as FORMAN+1. The first drawback associated with an aspatial model is that it lacks any consideration of the location of timber a key component in determining the economic feasibility of extraction (Baskent, 1990). The sustainable wood supply projected by an aspatial model is therefore an overestimate. Second, the schedule produced exists in tabular format as aggregates of stand types. The aggregates must be broken out into component stands so that harvest block mapping can occur. The result of this process is a map featuring stands with like harvest periods spread over the extent of the forest. Blocking of these stands is a long and tedious manual activity. GISFORMAN, on the other hand, provides a map-based harvest schedule that is a closer approximation to operational reality. Therefore, the forest manager is provided with a better starting point from which to begin the manual blocking process.

Gisforman, however, still lacks consideration of many operational realities of timber harvesting. Important criteria for block placement, such as proximity to existing road networks and mills, seasonal harvest distribution, product mix, and timber flow, are ignored by the model. Furthermore, the pseudo-blocks formed by GISFORMAN follow stand boundaries. A large portion of the manual blocking exercise involves changing harvest periods and "squaring-up" stand boundaries to make operational blocks. Even though the pseudo-block map produced by GISFORMAN provides a better starting point for the manual blocking process, there is still a great deal of editing required.

## Establishing the missing link

In order to improve upon the existing management planning process, a feedback loop has been created that provides a link between strategic and operations considerations (Figure 2). To accomplish this, a new routine has been created and some modifications made to the GISFORMAN model. The new routine, BLOCKER, will permit forest planners to digitize harvest blocks and assign harvest periods as part of the operational editing of model harvest schedules. The model was modified to include two new harvest scheduling options: pre-blocking, the default option, whereby only the block boundaries are used, and pre-scheduling, where harvest periods, as well as block boundaries, are used. The model can now be instructed to cut harvest blocks exactly as digitized and either determine the harvest periods based upon a user-supplied strategy (pre-blocking) or schedule blocks according to user-assigned periods (pre-scheduling). It is expected that forest planners will employ a mix of the two options.

Effectively, the forest landbase is netted down to the area held within the blocks digitized, in a fashion similar to the netting down of the landbase for wildlife concerns and riparian buffers. The end result is that forest response statistics normally provided on a pseudo-block basis ( growing stock volume, volumes, areas cut, planted, and thinned, etc.) can be developed for the digitized harvest block schedule.

Although a manual blocking step is still required, the improved planning process addresses several known problems. First, it will be easier (faster) to perform manual blocking with GISFORMAN's pseudo-block

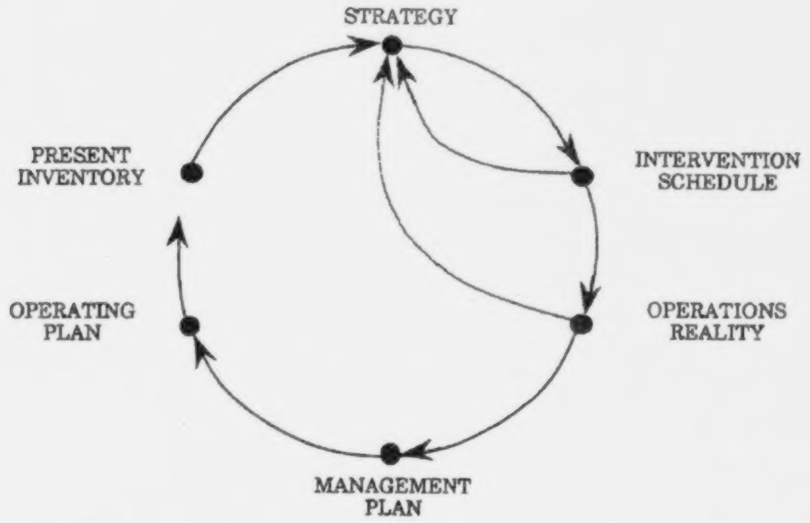


**Figure 2:** *The modified forest management process with a feedback loop between strategic and operations considerations.*

harvest schedules. Second, an accurate measure of harvest level reduction due to operational blocking is determined. Third, the wherewithal now exists in the planning process to develop and analyze alternative harvest blocking patterns. Fourth, the final harvest schedule is in digital format, permitting easy updating. Such common occurrences as destruction of stands by fire and insects, and removal of blocks from the schedule due to changes in policy, can be quickly made to the digital version. New maps showing the changes can be created just as easily.

Specific steps in the new management planning process are illustrated in Figure 3 and itemized below.

1. Strategic planning

   Run GISFORMAN 3.0 (latest release of GISFORMAN that incorporates all functionality of the FORMAN+1 model and the modifications described previously) to test strategies in a spatial context, and select one.

2. Harvest blocking

   Manually re-block the mapped schedule of pseudo-blocks created for the selected strategy to incorporate operationally important information (e.g. inventory data gathered from ground and aerial reconnaissance and personal knowledge of the landbase — i.e. any supplemental information that is inaccurate or missing from the provincial digital database).

3. Digitizing harvest blocks

   Digitize the harvest block schedule created in step 2 and assign harvest periods if the pre-scheduling option is chosen (period assignment is not required for the pre-blocking option). Process the information through the BLOCKER routine to create a digital, harvest block schedule.

4. Rerunning GISFORMAN

   Using either the pre-blocking or pre-scheduling harvest schedule option, rerun GISFORMAN with the schedule produced in step 3 and generate statistics regarding the new harvest level and its sustainability.

5. Iterating

   Repeat steps 2 through 4 if the sustainable harvest determined in step 4 is unacceptable.

## Making it work

Crown Forest Licensees in New Brunswick are currently in the process of preparing 25-year management plans, due April 30, 1992. Three Licensees J.D. Irving Ltd., Fraser Inc., and Juniper Lumber Co. Ltd will make use of the GISFORMAN model, and the modified planning process described previously, to assist in assessing harvest and silviculture strategies and developing associated schedules.

Juniper Lumber Co. Ltd will serve as the principal test case. The company is situated in the west-central portion of the province and holds the smallest of the 10 Crown Forest Licenses, with a landbase of 135000 hectares. Annually, 30-40 harvest blocks of various sizes are harvested to attain the annual allowable cut of 165 000 m3 of softwood. Juniper Lumber Co. Ltd does not possess the computer equipment required to run GISFORMAN and BLOCKER software. The facilities

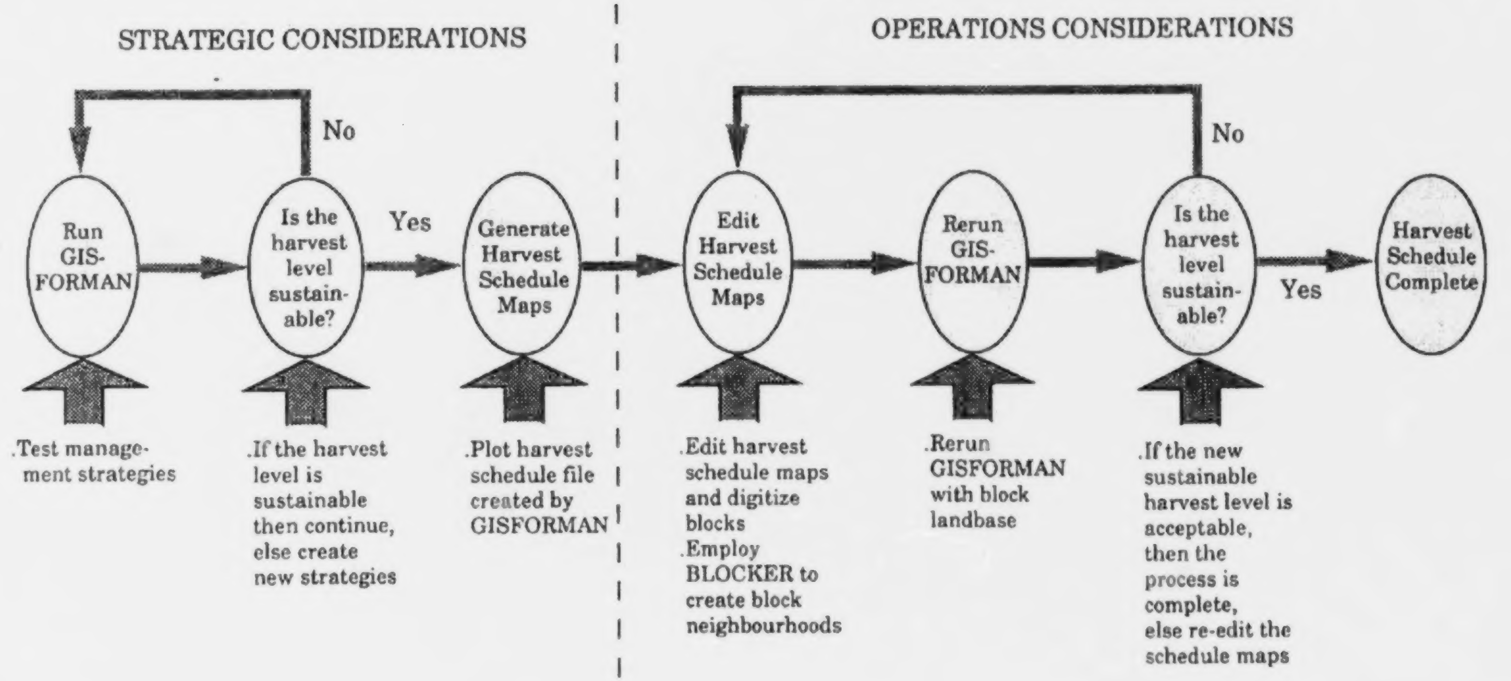

STRATEGIC CONSIDERATIONS
OPERATIONS CONSIDERATIONS
Run GIS-FORMAN
Is the harvest level sustainable?
No
Yes
Generate Harvest Schedule Maps
Edit Harvest Schedule Maps
Rerun GIS-FORMAN
Is the harvest level sustainable?
No
Yes
Harvest Schedule Complete
.Test management strategies
.If the harvest level is sustainable then continue, else create new strategies
.Plot harvest schedule file created by GISFORMAN
.Edit harvest schedule maps and digitize blocks
.Employ BLOCKER to create block neighbourhoods
.Rerun GISFORMAN with block landbase
.If the new sustainable harvest level is acceptable, then the process is complete, else re-edit the schedule maps

**Figure 3:** *Modified harvest scheduling process showing the iterative use of GISFORMAN and the link between strategic and operations considerations.*

| HARVEST / SILVICULTURE STRATEGY | BLOCK SIZE | BLOCK SHAPE | PERIOD DEVIATION | BLOCK PATTERN | BLOCK DISPERSION |
|---|---|---|---|---|---|
| | (small or large) | (regular or irregular) | (1 - n) | (regular or irregular) | (clustered or dispersed) |
| Minimize volume loss<br>Maximize harvested volume<br>Maximize product %<br>Minimize harvest cost<br>Minimize harvest block road cost<br>Maximize volume concentration<br>Minimize volume loss on a block basis | **Blocking factor effects on sustainable harvest level for each harvest and silviculture strategy.** | | | | |

**Figure 4:** *Proposed testing to determine the effect of blocking factors on sustainable harvest levels associated with harvest/silviculture strategies.*

at the Department of Forest Resources of the University of New Brunswick will be employed.

Participation in the actual 1992 planning process with Juniper Lumber Co. Ltd. will permit assessment of the new planning process. First, the functionality of new software will be tested and repaired if necessary. Second, the efficiency of the new process will be documented. In 1987, three employees of Juniper Lumber Co. Ltd. spent 220-250 person-hours over the course of 4-6 weeks on the management planning process. It is expected that due to an improved starting position (pseudo-blocks instead of scattered stands), the amount of time spent on blocking will be reduced considerably in 1992. Third, the pre-blocking and pre-scheduling harvest options of GISFORMAN will be compared. Allowing the model to assign harvest periods to digitized blocks (pre-blocking) has the advantage of maintaining a closer link to the original harvest schedule (and strategy) and its associated harvest level. As a result, it is expected that pre-blocking, without pre-scheduling, will have a lesser impact on sustainable harvest level reduction. Further, it should also be a time saver, since manually assigning harvest periods to blocks (pre-scheduling) is a rather complicated process due the need to satisfy adjacency limits (a two period (10-year) delay in cutting is mandatory for adjacent blocks in New Brunswick).

Although a mechanism has been created for linking strategic and operations considerations in management planning, its doubtful that Licensees will be able to take advantage of it in 1992 management planning. The stumbling block is a lack of cause-effect relationships. Foresters won't know how to change a blocking pattern across a whole forest to improve sustainable harvest level fall down. Basically, they lack a collection of blocking strategies comparable to harvest and silviculture strategies. A blocking strategy would be comprised of the best combination of block size, block shape, period deviation (from original model assignment), block dispersion, block pattern, and perhaps other, as yet undetermined blocking factors, that yields a certain sustainable harvest level for a particular harvest and silviculture strategy.

A study (experiment) is planned to establish some fundamental cause-effect relationships between the manner in which operational harvest blocking is done and sustainable harvest level reductions. The measurement of the blocking factors will be made in several ways. Block size will be varied from a minimum size to a maximum size. Block shape will be varied from a regular shape (circle) to a completely irregular shape. Shape will be measured as a function of perimeter to area in determining the degree of irregularity. Period deviation will be varied from 1 period of deviation from the model assignment to n periods. Block pattern will be varied from a regular arrangement of blocks (checkerboard) to a an irregular arrangement. Block dispersion will be varied likewise from clustered to widely dispersed. In order to develop blocking strategies, testing of various combinations of blocking factors is planned. Figure 4 details a proposed testing procedure that should reveal the effect of various factors on harvest level reduction and interaction with harvest/silviculture strategies. The first step will vary one factor while holding all others constant. Subsequent testing will involve varying combinations of factors and/or eliminating factors that have no measurable effect.

## Summary

The need for incorporating operational blocking considerations into the strategic level of the forest management planning process is obvious. The current New Brunswick approach of building elaborate schedules by incorporating extensive details about the dynamics of the forest is certainly taking the forest management process in the right direction. However, ignoring the impact of operational realities during the early planning stages is inefficient at best. A great deal of time is currently spent attempting to manually block the schedules produced by an aspatial modelling approach, but very little effort is exerted in determining the negative impact of operational modifications on initial strategies and its cause(s).

A BLOCKER routine and new harvesting options for the GISFORMAN spatial wood supply model have been developed to establish a loop between the strategic and operations considerations in the management planning process. While the loop in the process is currently not practical to execute due to a lack of cause-effect connections, a study is proposed to fill the void. It is expected that the study will permit some key factors contributing to harvest level reductions following operational blocking to be identified or confirmed. The payoff will be the capability to create and test, in a knowledgeable way, alternative harvest blocking patterns in the management planning process.

## Literature cited

Baskent, E. Z. 1990. *Spatial Wood Supply Modelling: Concept and Practice.* M.Sc.F. Thesis, University of New Brunswick, Fredericton, N.B. 85 pp

Baskerville, G. L. 1986. Understanding Forest Management. In: Dick, R.C, and G.A. Jordan, ed., *Forest Management and Geographic Information Systems*, Volume 1 - Beginnings, 1990, pp 8 - 22.

Erdle, T.A. and D.C. Frame. 1991. Timber Management in New Brunswick. In : Brand, D.G., ed. Canada's Timber Resources. Petawawa National Forestry Institute, Information Report PI-X-101, pp 87-95.

Ker, J. W. 1982. Evolution of Forest Management Policy in New Brunswick. *U.N.B. Forestry Focus*, Vol. 7, No. 2, Spring 1982.

Wang, E. C., T. A. Erdle and T. Roussel 1987. *FORMAN wood supply model user manual. Version 2.0,* NBFRAC, Fredericton, N. B., pp 60.

Vanguard Forest Management Services Ltd. 1991. *FORMAN+1 wood supply model user manual.* Version 1.0, Fredericton, N. B., pp 64.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Up the GIS learning curve: The New Brunswick forest management experience

**Robert C. Dick**

Director
Timber Management Planning
New Brunswick Department of Natural Resources and Energy
Fredericton, N.B.

## Abstract

*Forest management planning in New Brunswick is a given. Since the adoption of the Crown Lands and Forest Act in 1980 Crown land licensees have been required to prepare detailed plans that project long term wood supplies. Like most other skills however, perfecting the planning process takes time. This paper chronicles forest management planning in New Brunswick through three iterations from 1982 to 1992. It describes the increasing sophistication of the planning effort in a number of areas including use of GIS data; yield curve development and implementation tools. It concludes with the vision that future planning efforts will require sophisticated tools that take the drudgery out of data compilation but at the same time maintaining key user input.*

## Introduction

Forest management has been a reality in New Brunswick for a full decade. The catalyst for this effort has been the 1980 Crown Lands and Forests Act (CLFA) that put responsibility for management in the hands of Crown Licensees. The success of forest management in New Brunswick has been well documented by others (Jordan & Erdle, 1990; Erdle & Frame, 1991) and is due, in no small way, to the use of a geographic information system (GIS). Under the Act, planning schedules require that management plans be revised on a 5-year cycle. The third iteration of this process will take place in 1992 and it is perhaps useful to reflect on the process — to see where it has been, where it is now and where it is going.

This story cannot be told without including discussion of developments in wood supply and growth and yield modelling techniques. These discussions will be, however, kept to the minimum and this paper will highlight where GIS has aided the New Brunswick forest management planning process and brought it forward in significant ways.

## Origins of forest management in New Brunswick

As far back as 1971, New Brunswick became concerned that there were severe problems with its forest resource. In that year the Forest Resources Study (Anon, 1974) was initiated and delivered its final report in 1974. That Study recognized that there was no available surplus of the primary softwood species used in the Province and that there was an immediate need to take a proactive role in managing the forests to realize their full potential. The Study spun-off a short-lived Pilot Project in which organizational and management strategies were tested that eventually evolved into the Crown Lands and Forest Act.

The Act was adopted in 1980 and provided for the division of provincial Crown forest lands into 10 timber licenses. Each was assigned to 10 of the larger indigenous forest-based companies in the Province, in return for which each assumed specific obligations as stipulated in the forest management agreement and in accordance with the Crown Lands and Forests Act and its regulations.

These obligations, among others, involve the development and periodic revision by the Licensee, subject to the approval of the Minister of the Department of Natural Resources and Energy (DNRE), of three types of plans — the Industrial Plan, the Management Plan and the Operating Plan. The Industrial Plan describes the Licensees infrastructure and projects planned modifications to their business objectives for a 25-year period. While an important document, it is relatively easy to prepare and is developed by senior managers who have knowledge of long term company objectives. The Operating plan is for a one year period and details the specific forest management initiatives to be implemented on the Licensee for that year. Because it is drawn from the Management Plan, it also is relatively easy to prepare and is done so on a routine basis.

The Management Plan, is perhaps the most onerous of the planning documents required of Licensees. It is required to:

(i) describe development types within the forest on the basis of physical and biological differences;

(ii) estimate stand growth rates for these development types;

(iii) specify the level and types of silvicultural methods to be employed; and,

(iv) provide detailed mapping of the locations and quantities of product to be harvested by operating period and indicate the location and extent of silvicultural activity with respect to planting and thinning.

While the plan is revised on a 5-year cycle, it establishes sustainable wood volumes over an 80-year horizon and maps harvest and silviculture operations for 25-years.

In order to make this level of management planning possible it was recognized early on that there would have to be detailed inventory information available and that it would have to handled in an efficient manner. Almost two decades after they were written, it is interesting to reflect on the words from the 1974 Final Report of the Forest Resource Study that tried to enunciate this vision. The authors knew that New Brunswick had good inventory statistics on what timber resources there were in the Province but could see that they were very weak on "where timber of certain characteristics can be found". To fill this void they called for "forest cover maps of the entire Province to be produced as soon as is practicable" and went on to say that "all major disturbances of forest cover (cutover areas, recent burns, new roads or other structures, areas of forest destruction by any natural or man-made causes) be photographed from the air every year". They concluded by saying "the more information that is gathered, the more difficult it is to store it and to make it readily available for practical uses. We therefore strongly favor the establishment of a comprehensive land-data-bank".

It would take a number of years for the "land-data-bank" concept to become the geographic information systems that we know today but undeniably these early visionaries were on the right track. By the time the CLFA was adopted in 1980 there was a concerted effort on part of the DNRE to acquire such technology. The urgency for this acquisition was noted in correspondence of one senior DNRE official in 1981 when he said "not only must DNRE have a GIS to implement the intent of the Act, but it must have the system in place and fully operating 1.5 years before the management plan renewals in 1987. Without a GIS it is not possible to have planned scheduling of the harvest, to plan the product distribution, to plan silviculture or to efficiently protect the forest (and) evaluation of performance is silly without an information base" (Anon, 1982).

The Arc/Info GIS was eventually acquired in 1982 and a program of database loading began in earnest. It was in that same year that the first management plans were created. This was followed by revisions in 1987 and 1992. The balance of this paper will look at the development of those management plans in each of these years and will do so by focusing on four main areas of the planning process:

- Land Base Description
- Stand Growth Forecasting
- Forest-level Projections
- Spatial Configuration

## The first attempt at management planning in New Brunswick — 1982

With the fundamentals of the CLFA in place and boundaries of the 10 Licenses established, the stage was set for development of the first management plans. But few of the tools were in place.

### Land base description

The only complete inventory for the Province was created in the mid to late 1970s and was not designed to provide the types of data required for wood supply analysis. Regardless, the forest was described using coarse aggregations primarily based on broad covertypes. No site specific reductions were made to the land base to account for harvest restrictions. Instead an arbitrary 15% was removed from the entire forest across all classes.

### Stand growth forecasting

Because the existing inventory did not contain the right types of information that would allow realistic modelling of stand dynamics, descriptions had to be improvised. For example:

(i) there was little or no data in the inventory that would allow creation of yield curves to describe stand development. In lieu of such data, yield curves were basically drawn freehand using field foresters' knowledge of the forest.

(ii) there was no age information in the inventory that would allow placement of stands on yield curves. However, because volume data was available on a stand by stand basis in the existing inventory, a work around was developed whereby age was assumed to be linked to volume.

### Forest level projections

WOSFOP (Hall, 1977), an early vintage wood supply model, was used to forecast forest development and to determine sustainable harvest levels. It was lacking in several areas, e.g., it limited forest description to only ten classes.

### Spatial configuration

One of the major factors that the wood supply model did not take into account was the geographic distribution of stands. No formal attempt was made to disaggregate an accepted wood supply strategy into individual stands and resulting operating plans were mere best guesses at where actual wood volumes by stand types could be obtained.

As can be seen, the components needed for wood supply analysis were created in one fashion or another using whatever data and expertise that could be found. While it was Licensee responsibility to develop the actual plans, because of the limited expertise that was available in the Province, the actual wood supply runs were done by DNRE and the results given to the Licensee for inclusion in a formal document. Needless to say the effort was less than perfect but at the very least the program was underway and both Licensees and DNRE were learning. The need for GIS in developing management plans was never more evident and creation of a new inventory database that incorporated those descriptive data important to management planning became the focus of activity for the next several years.

## The first revision of management plans — 1987

The GIS-based forest inventory was officially completed in 1987. It covered the entire Province and encompassed approximately seven million hectares on 2000 map sheets defining some one million stands. It is now updated on a yearly basis for all harvesting on Crown land with periodic updates for small-private ownerships.

### Land base description

The new inventory was geared, in part, toward providing stand descriptions suitable for wood supply analysis. Because the inventory was fully in digital form on Arc/Info, stratification of the forest was simplified using the database capabilities of the GIS. However, its spatial analysis capabilities were utilized to only a limited extent. For example:

(i) accounting for the area in riparian zones was generally done manually using hardcopy map products. However, there were two Licensees in the Province who also now owned GIS technology and who did some limited buffer analysis using these systems.

(ii) there was a desire to include a measure of site quality in the planning process. This was conveniently accomplished, albeit at course resolution, by assigning a measure of site to each 1:12 500 map sheet. The GIS was then used to assign a site factor to each stand on each map sheet.

### Stand growth forecasting

To obtain a better measure of stand yields, a methodology for yield curve construction was developed that utilized the network of temporary and permanent plots that were established in support of the inventory. The methodology was not precise but gave users some hard data to support or reject yield curves developed in 1982.

### Forest level projections

Based on the short comings of the wood supply model used in the 1982 exercise, the intervening years were used to develop a new model called FORMAN (Wang et al, 1987). It was well tailored to New Brunswick forest conditions and made more realistic forest dynamic assumptions. It still however required aggregation of stands into classes.

### Spatial configuration

The lack of location-specific detail continued to flaw the process. In an attempt to compensate, the GIS was used to spatially allocate FORMAN's class-based harvest schedules and stand harvest maps were generated. It remained, however, to develop an operationally acceptable harvest block schedule from these maps — an arduous manual process. The important result was translation of a harvest schedule, designed at the management planning level, into a map format that had direct relevance at the operational level.

By the end of this second revision of management plans the process was much better understood by all parties and important strides had been taken to make planning assumptions better reflect reality. One major shortcoming however was that the process was almost fully geared to problems of sustaining fibre supply. Little or no accounting was made for multiple use, especially with regard to sustaining wildlife habitat. This item would be the focus of considerable development over the next 5 years.

## The current state of New Brunswick management planning

The management plans being developed for 1992 are the most sophisticated to date. Most notably, there are requirements that Licensees incorporate wildlife management in their plans by providing for deer habitat and habitat for fauna that frequent mature forest.

### Land base description

GIS was used extensively to more accurately describe the available land base. This was done by digitizing land base restrictions (i.e., restricted from a harvesting perspective) and overlaying them with the forest to report the area of covertypes within restrictions. These restrictions included:

(i) *Riparian Buffers:* To reflect reality, historical average buffer widths by watercourse type on a Licence-by-Licence basis, were determined and applied to all Crown watercourses within the Province.

(ii) *Deer Yards:* Identified Deer Wintering Area Management Units were digitized.

(iii) *Legal Reserves:* Ecological Areas and Sugar Bushes were digitized.

(iv) *Inaccessible Areas:* Topographic maps were used to identify areas of excessive slope that would not be harvested. These were subsequently digitized.

### Stand growth forecasting

For 1992, computerized growth models were developed that simulated the establishment, growth and decline phases of stand development using plot data from forest stands. These empirical models were not intended to give definitive yield projections but, as in 1987, were intended to guide users in their development of yields.

### Forest level projections

Primarily to accommodate specialized harvesting techniques required to enhance and maintain wildlife habitat, the FORMAN wood supply model was upgraded (FORMAN+1 and FORMAN+2, Vanguard, 1991) to allow for 2-pass, shelterwood and true uneven-aged management.

### Spatial configuration

For the majority of Licensees, the GIS-based approach used to develop 1987 harvest schedule maps will be used again in 1992. Three Licensees, however — J.D. Irving Ltd., Fraser Inc. and Juniper Lumber Ltd. — will employ a spatial wood supply model. These companies are participating in the FORMAN 2000 project at the University of New Brunswick and will use the GISFORMAN model in developing and testing wood supply strategies. As the name implies, GISFORMAN is an extension of the original FORMAN but differs in two important ways:

(i) the model produces forest forecasts based upon stand-level information as opposed to stand aggregate information; and,

(ii) by accessing a topologically structured database, the model mimicks harvest blocking and adjacency delay in developing and testing harvest strategies.

## The future of management planning in New Brunswick

Table 1 summarizes the development trends that have been discussed above. Clearly there has been a trend toward sophistication of the procedure, particularily in terms of:

(i) more accurate descriptions of the nature of stand growth and yield;

(ii) incorporation of multiple use into the planning process; and,

(iii) maintenance in the planning effort of the spatial characteristics of the forest.

These trends are likely to continue and be evident in the 1997 planing effort. Because of its cyclical nature there has been a tendency for the stakeholders to focus on the problem only once every 5 years. The increasing complexity has caused these stakeholders to realize that the process is important to the future well being of the Province and requires commitment of staff and resources on a continual basis to ensure that plans adequately reflect the goals and objectives of forest management.

Consistent with this trend, will be a shift from DNRE leadership in planning to leadership by Licensees. This shift is healthy and consistent with the primary intent of the CLFA, i.e. that licensees be responsible for management and that DNRE be in the position of setting objectives and monitoring performance.

New Brunswick has been fortunate in that this level of forest management planning experience has attracted considerable expertise to the Province. Not only is there considerable interest in research on the subject at the university level but there are now several consulting firms in the Province that have developed marketable skills in the area of GIS and wood supply modelling.

| Activity | 1982 | 1987 | 1992 |
|---|---|---|---|
| **LandBase Description** | • old inventory<br>• gross removal of 15%<br>• no other reductions | • new inventory<br>• more realistic reductions (manual)<br>• limited use of GIS | • inventory updated for removals<br>• land base reduced using GIS |
| **Stand Growth Forecasting** | • "best guess" used<br>• limited plot data | • methodology developed using plot data | • stand yield models developed |
| **Forest-Level Projections** | • simplistic model (WOSFOP) | • more realistic model (FORMAN) | • FORMAN expanded for special harvest techniques<br>• GISFORMAN allows true spatial modelling |
| **Spatial Configuration** | • poor link between strategy and operational plans | • harvest schedule developed from FORMAN using GIS<br>• manual blocking | • general Lic. use of 1987procedure<br>• 3 Licensees develop blocked schedules from GISFORMAN |

**Table 1: Trends in forest management planning in New Brunswick between 1982 and 1992. (adapted from Erdle and Frame, 1991)**

At the heart of the planning exercise is an up to date forest inventory. While the New Brunswick inventory has been updated on an annual basis for cuts and burns, the unharvested inventory will be more than 10 years old by the time the next management plan revisions are due in 1997. Consequently innovative procedures are currently being investigated to determine ways of updating this inventory and at the same time maintaining the investment in the updated information in the existing database.

## Lessons learned

The New Brunswick experience is neither unique nor perfect. It has demonstrated, however, that management planning on a forest-wide basis can be taken beyond a theoretical exercise and the lessons learned to date may prove useful to others contemplating similar initiatives:

(i) Nothing as complex as the New Brunswick forest management planning process could be implemented error-free. However, waiting for perfection only results in inaction. This has been an underlying, though not specifically stated principle driving the New Brunswick forest management planning effort. As Samuel Johnson said "nothing will ever be attempted if all possible objections must first be overcome".

(ii) The success of the Arc/Info geographic information system at DNRE is due in large part to the Crown Lands and Forest Act. Had it not been for the legislated mandate of the Province to provide Licensee managers with up to date information on the Crown forest for planning purposes, it is doubtful that the system would be in use today.

(iii) The process is complex and constantly changing. Not only do standards change but goals, objectives, rules and guidelines change. Couple this with changes in staff and soon there is a process understood by few and mistrusted by many. To avoid this situation there is a need to simplify the complexity without taking away the need for input by the forest managers. The first step in this process is thorough documentation of the procedure followed by the possible development of knowledge-based computerized tools that prompt managers for input and guide them through strategy development.

## References

Anon. 1974. Report of the Forest Resources Study. Province of New Brunswick. 362p.

Anon (unpublished manuscript). 1982. The Evolution of the Crown Lands and Forest Act. Province of New Brunswick, 49 p.

Erdle, T.A. and D.C. Frame. 1991. Timber Management in New Brunswick. In: Brand, D.G., ed. Canada's Timber Resources. Petawawa National Forestry Institute, Information Report PI-X-101, p. 87-95.

Jordan, G.A. and T.A. Erdle. 1990. GIS and Forest Management: What Have We Learned in New Brunswick. CISM Journal, Vol. 43, No. 3, p. 287-295.

# B4 Fire applications

**Tracking suppression resources via an Australian fire management system: in principle and in practice**

Stephen Kessell et al., Curtin University of Technology, Perth, Western Australia

**Evaluating spatial strategies of wildfire prevention with a GIS**

Chou Yue Hong, University of California, Riverside, CA, USA

**The application of GIS technology for forest fire management planning**

Bryan Lee et al., Forestry Canada, Edmonton, AB, Canada

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Tracking suppression resources via an Australian fire management system: in principle and in practice

**Judith A. Beck and Stephen R. Kessell**

Department of Geographic Information Systems
School of Computing
Curtin University of Technology
GPO Box U1987
Perth, Western Australia 6001

## Abstract

*A geographic information and modelling system has been developed in Western Australia to support the management of logistics during wildfire suppression, concomitantly with simulation routines that support an appreciation of fire behaviour. This work details the design, structure and use of generic routines that support the tasks of planning and tracking suppression resources, giving due consideration to critical financial, managerial and organisational constraints which have governed the development of this system.*

*Textual information that details a particular suppression resource is associated with an icon that categorises it visually. Mobile resources can be positioned at their work location with the click of a mouse. The provision of a simple, visual summary is interpreted readily, and functions as a valuable incident briefing tool.*

*The utility of this system, which operates on personal computers, lies in its ability to integrate fire behaviour, suppression, plant, personnel, environmental, historical and cultural features in a manner useful to fire controllers.*

## Introduction

A Wildfire Incident Management System (WIMS) is being developed to support decision making during wildfire control. The system, the inception of which is detailed elsewhere (Beck 1989, 1990, 1991; Kessell 1990; Kessell and Beck 1991), can be conceptualised via its two primary functional components: fire spread routines are used to simulate the spatial development of a forest fire over time; and resource management routines are used to deploy, track and query suppression resources for a particular event.

The development of decision support systems for fire management must consider the special objectives of the organisation at hand, and the development of this system has been governed by the financial, managerial and organisational constraints of the Western Australia Department of Conservation and Land Management (CALM). During the Cyclone Alby emergency in 1978, when 92 fires developed in State forest over a three hour period (Underwood *et al.* 1985), it was extremely difficult to consider all values at risk, to identify suppression priorities and to track the whereabouts of allocated resources. With the development of WIMS, CALM seeks to improve the safety, efficiency and effectiveness of suppression efforts under extreme circumstances such as those presented by Cyclone Alby.

The majority of WIMS is PC based, because phone, electricity and main-frame computer operations cannot be guaranteed during adverse situations in Western Australia. Hence, each installation must be self sufficient with respect to data and software.

In order to ease the transition from a manual to an automated decision support system, design strategies have concentrated on incorporating concepts from as many of CALM's manual systems as possible. For example, the Western Australian forest fire behaviour tables (Sneeuwjagt and Peet 1985) provide the basis for fire behaviour predictions. A fire growth model, which is a hybrid between cell and vector based perimeter spread algorithms such as those developed by Kourtz and O'Regan (1971) and Anderson *et al.* (1982), employs weather inputs, geographic information and the Western Australian fire behaviour tables to predict the temporal and spatial development of a fire.

The resource allocation and tracking routines of WIMS have incorporated concepts from two of the manual systems that CALM uses for logistics management. These two manual systems are namely: the Fire Resource Management System (Hamilton 1988, Rawet 1990), which employs colour coded cards to depict allocated resources such as personnel, plant and equipment; and, the Situation Board, which is used to map the physical infrastructure and depict the whereabouts of deployed resources with colour coded magnets.

The Department of CALM has also developed a computerised Resource Inventory System, which is known as RESIN (Rawet 1991). RESIN is a simple, data base management application, which has been developed in DBase, that is used to maintain current listings of available fire control personnel, plant, vehicle and equipment resources. Hence, this system could be used as a basis for resource management in WIMS.

The dynamic allocation routines of WIMS are used to track the whereabouts of resources that have been deployed for a particular event. These modules can be integrated with any data base management applications that has been programmed in DBase or Clipper. Textual information that details a particular vehicle, plant or piece of equipment can be associated with an icon that categorises it visually. Mobile resources such as tankers, bulldozers and aircraft can be positioned at their work location with the click of a mouse. Plant, vehicle and equipment resources can be affiliated with the individuals responsible for them, and icons can be interrogated for personnel or equipment details.

## User interface

WIMS does not incorporate digitising or Boolean map algebra routines of a typical GIS. An emergency situation is hardly the time to begin to collate forest type, fuel quantity, cadastral and cultural details. Instead, WIMS is expected to compliment a corporate GIS, which functions to capture, maintain and manage the geographic information that is relevant to fire management. Facilities have been developed to transfer geographic information between WIMS and commercial GISs such as ARC/INFO or INTERGRAPH.

WIMS has been designed to incorporate whatever geographic information and suppression resource details are required in order to meet the needs of a given organisation. In effect there are two operating levels of WIMS. Low level system manager routines are used to define the structure of geographic and suppression resource information for a particular installation. The general user, on the other hand, is able to access this information as required, and uses the system to forecast, monitor and document the progress of events.

### System management level

System management routines are used to define, for example, the geographic information that is available, the mapping patterns that will be used to depict a particular theme value. These routines are also used to define the data bases that will contain plant, vehicle and other suppression resource details, as well as the icon images that will be used to depict a particular suppression resource.

This design framework has been provided so that ownership preferences can be incorporated dynamically. State headquarters can establish standard data bases, to which region or district installations can add. Hence, a degree of user ownership and responsibility can be established for each installation.

It is at the system management level that an icon is defined (Figure 1). The data structure of an icon is defined first, and its image is established subsequently.

Up to twelve characters are used to distinguish one icon from another by name. The mobility of an icon is then established. A general user can place a mobile icon using a mouse to drag it to the desired location on the background map. Locational coordinates are then updated automatically. If an icon is static, however, the general user cannot alter its locational coordinates. For example, most suppression resources would be mobile icons, whereas point features such as rare and endangered flora and fauna would be static.

Next, the relevant data base file is selected from a point-and-shoot menu, which lists all of the available data base files. In Figure 1, for example, a data base (WWVEH.DBF) that contains vehicle information has been identified. In this case, all vehicle details are maintained in one data base file, and the system manager needs to distinguish between icon types on the basis of the vehicle type field (VTYPE). A record in the data base file is said to be a HEAVYDUTY icon if its vehicle type field is equal to "HD". Hence, the system manager can create different icons to depict a heavy duty, a bulldozer, a car, or a light patrol unit, for example.

The whereabouts of each icon is stored in two fields that define its x (Easting Field) and y (Northing Field) locational coordinates. These must be integer fields, and are intended to store Australian Map Grid (AMG) coordinates.

Equipment, plant and vehicles have people who are responsible for them. Hence, a HEAVYDUTY icon can be linked to a PASSENGER record, in the data base which contains staff details (WWSTAFF.DBF). It should be noted that locational coordinate fields are compulsory for an icon, but they are not required for an icon link. Further, several records of a designated link type can be affiliated with a particular icon. For example, a number of people can travel in a particular vehicle, and it's up to the general user to link a particular vehicle with its passenger details, interactively.

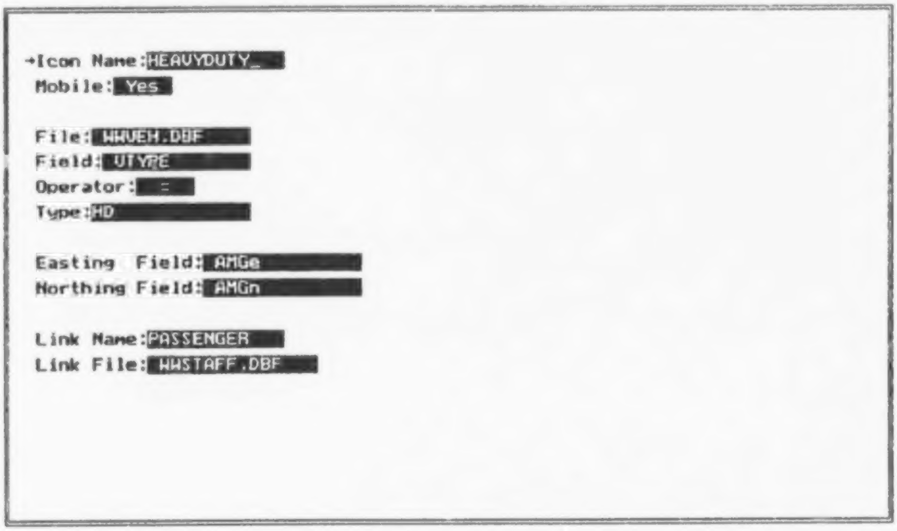
Next, the icon image is created. Up to 20 drawing elements such as circles, boxes, lines and triangles, are combined to define an icon's image. Elements are drawn in the active palette colour.

The set up for a given installation of WIMS need only be established once, and perhaps modified from time to time as required. Anyone using the system management routines must be familiar with the data base management application and its data structures, hence the system manager is expected to be relatively computer literate.

### General user environment

The general user need not be concerned with data base structures or GIS. In this environment, WIMS is used to create a visual situation report, an example of which is given in Figure 2. A base map, at the user's choice of scale, is used to depict the general surrounds of an incident: for example, district boundaries, roads and waterways can be superimposed on a map of any available data theme. In Figure 2, major and minor roads are depicted on an elevation map.

WIMS can be used to perform a number of tasks: for example, suppression resources can be allocated; a geographic location can be queried for attribute details; an icon can be queried for data base details; and search facilities can be used to locate a particular vehicle.



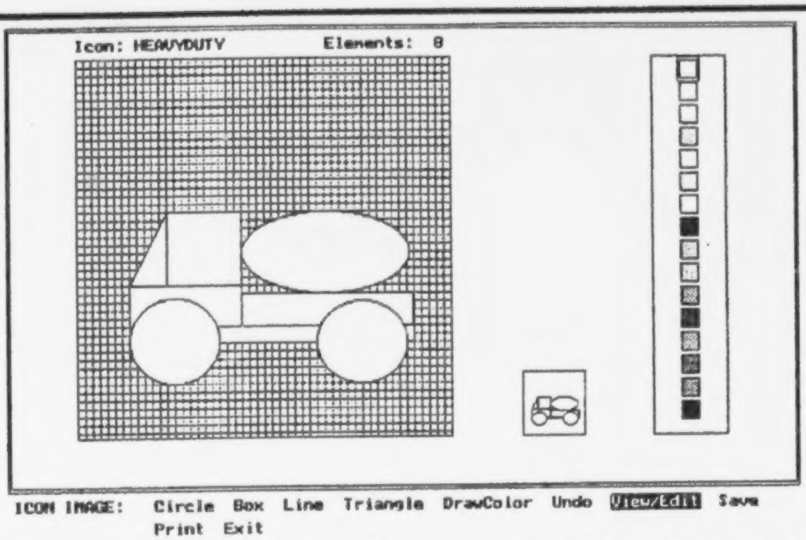
```
→Icon Name:HEAVYDUTY_
 Mobile: Yes

 File: WWVEH.DBF
 Field: VTYPE
 Operator: =
 Type:HD

 Easting  Field: AMGe
 Northing Field: AMGn

 Link Name:PASSENGER
 Link File: WWSTAFF.DBF
```

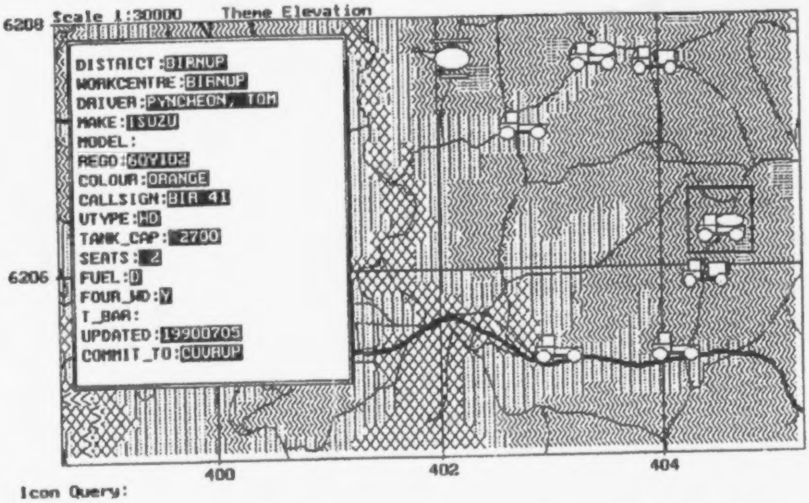
Figure 1a. The system management routines of WIMS are used to design the structure of an icon.



Icon: HEAVYDUTY     Elements: 8

ICON IMAGE:
Circle
Box
Line
Triangle
DrawColor
Undo
View/Edit
Save
Print
Exit

Figure 1b. The system management routines of WIMS are used to design the image that categorises it visually.



DISTRICT :BIRNUP
WORKCENTRE :BIRNUP
DRIVER :PYNCHEON, TOM
MAKE :ISUZU
MODEL :
REGO :60Y102
COLOUR :ORANGE
CALLSIGN :BIR 41
UTYPE :HD
TANK_CAP : 2700
SEATS : 2
FUEL :D
FOUR_WD :Y
T_BAR :
UPDATED :19900705
COMMIT_TO :CUVRUP

Icon Query:

Scale 1:30000
Theme Elevation
6208
6206
400
402
404

Figure 2. The general environment allows a user to track suppression resources, as they are allocated to a particular incident. The Query option can be used to examine the data affiliated with a particular icon.

Several steps are involved in tracking suppression resources via WIMS. The Allocate option is used select a particular icon type and record from a point-and-shoot menu, and drag it to its work location using the mouse. Links can then be created to affiliate a resource with the person(s) responsible for it, and the location of an allocated resource can be altered with the Move command. A user cannot deploy a resource twice, hence, if an icon or link record has been allocated, a user cannot re-deploy that resource until the icon has been recalled or the link has been broken. Icon queries present data base details, and a user can step through information that has been linked to an individual resource. A search filter can be used to identify those point features whose icon records or links adhere to user specified field criteria.

It should be noted that a record must exist in the data base if it is to become an icon or link. WIMS has not been designed to allow for the addition, deletion, or modification of a data base. One of the primary functions of a good data base management application is to ensure that the data is checked on the way in; WIMS would have to incorporate an infinite number of tailoring facilities at the system management level in order to provide suitable data management routines from within. Similarly, resource reporting facilities would have to cope with providing endless report format options.

Data base management and report writing tasks are more appropriately carried out by the data base application, where they can be tailored to meet the needs of a particular organisation. Instead, WIMS allows a user to invoke a number of data base management applications from within. In fact, any executable program can be run from WIMS, provided there is sufficient memory available to load and execute the child process.

## Discussion

In practice, the suppression tracking routines of WIMS are not apt to replace the current manual T-board systems, such as the Fire Resource Management System (Sneeuwjagt and Vear 1987; Hamilton 1988; Rawet 1990), that are used for logistics planning. Establishing icon and data relationships interactively may be tedious, although key stroke requirements are reduced by applying point-and-shoot techniques, which allow a user to select relevant details from an established list. Details that are not available in the local data base have to be keyed in anew, and this could prove to be rather time consuming during events that involve substantial resources from external agencies.

One alternative to positioning individual resources manually, would be to update their locations automatically using Global Positioning Systems. This is not apt to be economically viable for the Department of CALM. Such a tact might be cost-effective if the use of aircraft in remote suppression operations were more prevalent than they are currently in Western Australia.

The value of the suppression resource tracking component of WIMS lies in its ability to integrate any local data base readily. Static, intelligent point features can be used to locate and detail rare and endangered flora and fauna, archaeological sites, suppression hazards or any other feature that merits due consideration during suppression planning. In Australia, the ecological impacts of fire suppression activities are of increasing concern, and this type of information will only be considered in suppression planning if it is available readily.

There are many reasons why WIMS has been implemented on a cheap, portable, stand alone system platform. Personal computers no longer pose the speed and memory limitations they once did, and they are not dependant on, but can work in conjunction with, main-frame systems.

The is a more important reason for maintaining a relatively low level of technology in geographic information, modelling and decision support systems such as WIMS. Although the capabilities of GIS implemented on sophisticated graphics workstations helps communicate information, the "pretty pictures" generated by the technology may present a false lure about the credibility of the information depicted (Burrough 1986, 1990).

This is especially relevant in the case of modelling the temporal and spatial development of a fire, where modelling and data weaknesses are apt to be prevalent. The mathematical modelling techniques that are used to simulate fire spread (see, for example, Kourtz and O'Regan 1971; Anderson *et al.* 1982) have the potential to introduce bias or error, which is a direct result of the simulation process and is distinct from that associated with input parameters (O'Regan *et al.* 1976; Richards 1990; Feunekes 1991).

Rather than focus on the technology *per se*, the introduction of WIMS in March 1992 will stress limitations in the prediction models and data. If simulation outcomes are to be applied effectively, a "license to drive" can only be granted to user's who have a sound knowledge of fire behaviour. For these user's, WIMS will foster their knowledge, as well as structure and support the decision making processes critical to wildfire management.

## Literature cited

Anderson, D.H., E.A. Catchpole, N.J. DeMestre and T. Parkes (1982) Modelling the Spread of Grassland Fires. J. Austral. Math. Soc. (Series B), 13, 452-466.

Beck, J.A. (1989) A Geographic Information and Modelling System for the Management of Wildfire Incidents. *In* Proceedings of the 17th Australasian Urban and Regional Planning Information Systems Conference, Perth, Western Australia.

Beck, J.A. (1990) The Development and Application of Generic Map Algebra and Modelling Tools. *In* Proceedings of the 18th Australasian Urban and Regional Planning Information Systems Conference, Canberra, Australian Capital Territory.

Beck, J.A. (1991) GIS and Modelling Technologies for Fire Research and Management in Western Australia. *In* Proceeding of the 11th Conference on Fire and Forest Meteorology, April 16-19, 1991, Missoula, Montana.

Burrough, P.A. (1986) Principles of Geographical Information Systems for Land Resources Assessment. Oxford: Clarendon Press.

Burrough, P.A. (1990) Preface: Methods of Spatial Analysis in GIS. International Journal of Geographical Information Systems. 4(3):221-223.

Feunekes, U. (1991) Error Analysis in Fire Simulation Models. MSc. Thesis, The University of New Brunswick, Department of Forest Resources, Fredericton. 75pp.

Hamilton, F. (1988) Fire Resource Management System. Fire Protection Branch, Department of Conservation, Forests and Lands, Melbourne, Victoria.

Kessell, S.R. (1990) An Australian Geographical Information and Modelling System for Natural Area Management. International Journal of Geographical Information Systems, 4(3):333-362.

Kessell, S.R. and J.A. Beck (1991) Development and Implementation of Forest Fire Modelling and Decision Support Systems in Australia. *In* Proceedings of GIS/LIS'91, Atlanta.

Kourtz, P.H. and W.G. O'Regan (1971) A Model for a Small Forest Fire... to Simulate Burned and Burning Areas for use in a Detection Model. Forest Sci. 17(2):163-169.

O'Regan, W.G., S. Nozaki, and N.E. Savin (1976) Bias in the Contagion Analog to Fire Spread. Forest Sci. 22(1):61-68.

Rawet, D. (1990) Fire Resource Management System. Fire Protection Branch, Department of Conservation and Land Management, Perth, Western Australia.

Rawet, D. (1991) RESource INventory System: User's Manual. Fire Protection Branch, Department of Conservation and Land Management, Bunbury, Western Australia.

Richards, G.D. (1991) An Elliptical Growth Model of Forest Fire Fronts and its Numerical Solution. Int. J. Numer. Methods Eng., 25, 625-634.

Sneeuwjagt, R.J. and G.B. Peet (1985) Forest Fire Behaviour Tables for Western Australia (3rd Edition). Department of Conservation and Land Management, Perth, Western Australia.

Sneeuwjagt, R.J. and K. Vear (1987) Fire Suppression Organization Booklet. Department of Conservation and Land Management, Perth, Western Australia.

Underwood R.J., R.J. Sneeuwjagt and H.G. Styles (1985) The Contribution of Prescribed Burning to Forest Fire Control in Western Australia: Case Studies. *In* Fire Ecology and Management in Western Australian Ecosystems (J.R. Ford, ed.). Proceedings of May 1985 Symposium, WAIT Environmental Studies Group Report No. 19, Perth, Western Australia.

GIS'92

Symposium

Vancouver
British Columbia
February 1992

# Evaluating spatial strategies of wildfire prevention with a GIS

**Yue Hong Chou**

Department of Earth Sciences
University of California
Riverside, CA 92521

## Abstract

*Effective management of wildland fires relies more on prevention than on suppression because preventive treatments cost much less than suppression measures. While suppression efforts demand immediate attention and quick-response actions, decisions on preventive treatments are harder to make unless their cost-effectiveness can be accurately evaluated. This paper presents the use of GIS to evaluate alternative spatial strategies of prescribed burning. The evaluation of a spatial strategy is based on two criteria: the fire danger of the entire district and the spatial pattern of fire occurrence probability. Eight spatial strategies of prescribed burning for the Idyllwild quadrangle of the San Bernardino National Forest were evaluated and the most cost-effective strategy identified.*

## Introduction

Wildfires are most controllable when they are small. Those fires that escape the initial attack of suppression forces cause substantial damage to natural resources and property, and, occasionally, loss of life. As a fire grows in size, both the damage it may cause and the cost to control it increase exponentially. Therefore, effective wildfire management requires a management plan based on which possible spatial strategies of preventive treatments can be evaluated. As critical localities of great fire danger are identified in the management plan, fire managers are able to make timely decisions on launching initial attack should a fire occur. More importantly, the management plan can be used to set the priority of possible proposals of preventive measures such as prescribed burning.

Prescribed burning is the application of fire to wildland fuels under certain conditions of weather, fuels, and topography that specific objectives can be accomplished safely (Green 1981). When carefully planned, prescribed burning can be much more cost-effective than suppression.

While technology in prescribed burning has been well developed (Emrick and Adams 1977, Green 1981), little has been done regarding the evaluation of spatial strategies. A spatial strategy is one that addresses the issue of location. In terms of prescribed burning, a typical spatial strategy is one that specifies the locations where prescribed burning will be applied. For managing a large district of wildlands under a limited budget, the capability ofevaluating spatial strategies is particularly important because preventive treatments must be planned effectively.

In previous studies (Chou *et al.* 1990, Chou 1992), GIS procedures were developed for constructing probability models of fire occurrence and, accordingto such models, delineating the distribution of critical zones of fire danger. This paper presents the empirical application

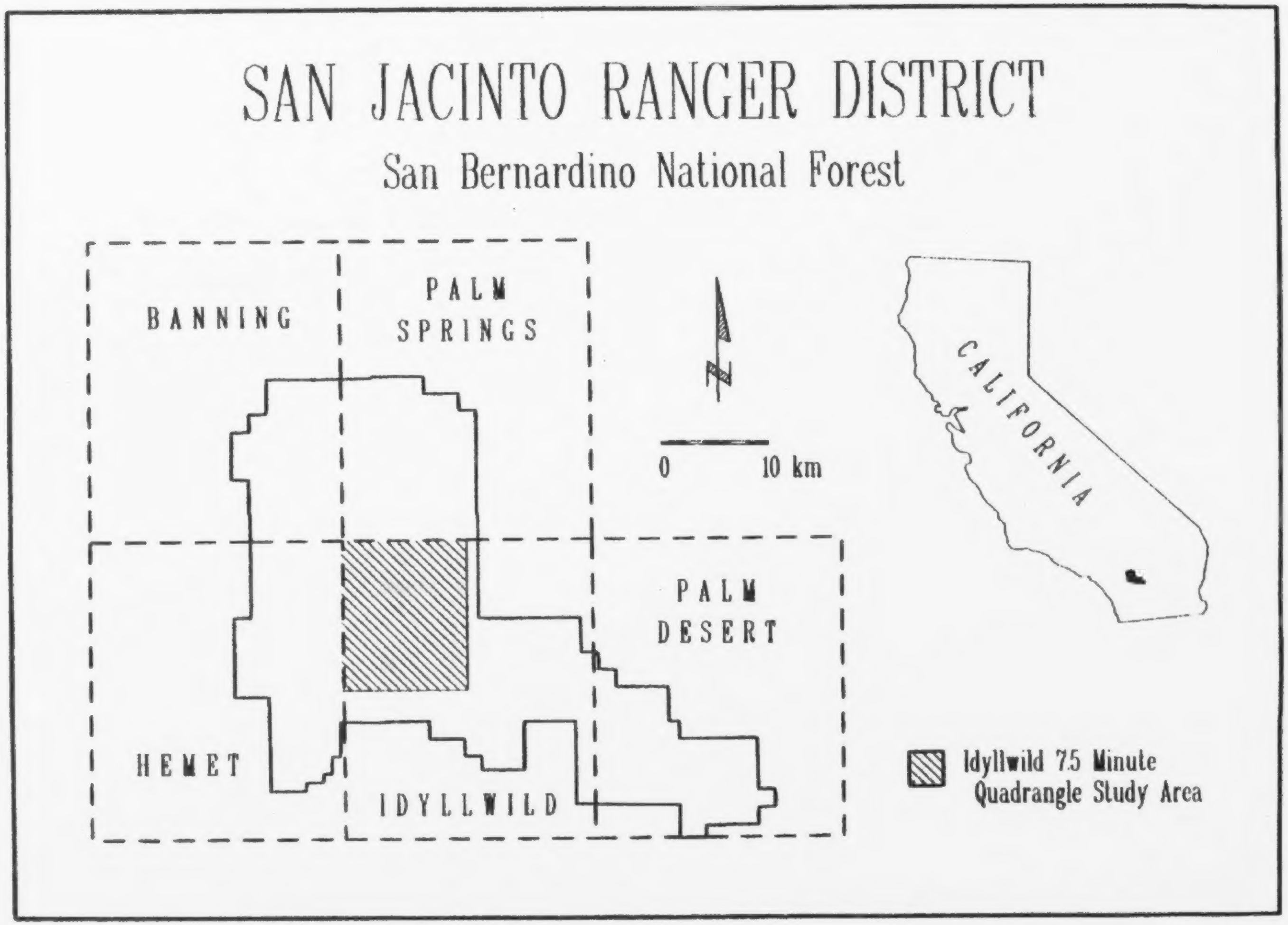

SAN JACINTO RANGER DISTRICT
San Bernardino National Forest
BANNING
PALM SPRINGS
HEMET
IDYLLWILD
PALM DESERT
0
10 km
CALIFORNIA
Idyllwild 7.5 Minute Quadrangle Study Area

Figure 1. The Idyllwild 7.5-minute Quadrangle, San Jacinto Ranger District, San Bernardino National Forest, California.

of those procedures to the evaluation of eight spatial strategies of prescribed burning in the Idyllwild 7.5 minute quadrangle of the San Bernardino National Forest, California (Figure 1).

## Setting spatial strategies

In the ecological database of the study area, fires that occurred between 1911 and 1984 were coded (Figure 2). Geographic units were defined by topographic characteristics such as slope aspect and slope gradient (Figure 3). Using the TIN package of ARC/INFO (Environmental Systems Research Institute 1987), the GIS employed in this study, the digital elevation models (DEM) of U.S. Geological Survey (1987) were converted into triangulated irregular networks which were then used for generating geographic units of homogeneous topographic characteristics. Figure 4 shows 200-meter contours generated from the triangulated irregular networks.

The distribution of fire occurrence probability is constructed from the logistic regression model (Wrigley 1976) such that

$$P_i = \frac{EXP(U_i)}{1 + EXP(U_i)} \tag{1}$$

where $P_i$ denotes the probability for a major fire to occur in the i-th geographic unit. Formally, the quantity U of the model is specified as:

$$\begin{aligned} U_i = {} & \beta_0 + \beta_1 AREA_i + \beta_2 ASPECT_i + \beta_3 SLOPE_i + \beta_4 ROAD_i + \beta_5 BUILD_i \\ & + \beta_6 ROTATION_i + \beta_7 RAIN_i + \beta_8 TEMPERATURE_i + e \end{aligned} \tag{2}$$

where $AREA_i$ is the area of the i-th geographic unit; ASPECT denotes slope aspect; SLOPE represents slope gradient; ROAD, the proximity to transportation, is the distance between the unit and its nearest road; BUILD, the proximity to man-built structures, is the distance between the unit and its nearest building; ROTATION is a fire rotation weight determined by vegetation type; RAIN is annual precipitation; TEMPERATURE denotes the mean temperature in July; ${}_j$ is the parameter for the j-th variable; e is the random error term. Coefficients of the parameters were estimated from the stepwise logistic regression program (LR) of the BMDP statistical package (BMDP 1987).

The probability of fire occurrence must be modified for bare ground and water surfaces. Since major fires are not expected to occur on these surfaces, the probability of fire occurrence should be assigned zero rather than being estimated from the model. For other surface types, the probability is derived directly from the best-fit logistic model. For management purposes, geographical units were classified into three categories by probability of fire occurrence: areas with a probability greater than .60 are labeled high fire-danger zones, those between .60 and .30 are labeled medium fire-danger zones, and less than .30 are labeled low fire-danger zones. Figure 5 shows the distribution of fire-danger zones based on this classification scheme. From this map, we delineated zones of high fire danger and, accordingly, defined eight spatial strategies of prescribed burning. In each strategy, three or four adjacent polygons will be burned. Figure 6 shows the polygons that need to be burned for the eight strategies listed in Table 1.

## Criteria for evaluation

The effectiveness of a spatial strategy of prescribed burning is evaluated by two measures: the district fire danger index (DFDI) and Moran's I coefficient of spatial autocorrelation.

The DFDI is the area-adjusted average of fire occurrence probability for the entire district, such that,

$$DFDI = \Sigma_i P_i A_i \,/\, \Sigma_i A_i \tag{3}$$

where $P_i$ is the probability of fire occurrence for the i-th geographic unit; $A_i$ is its area. In (3), the numerator is the sum of products of fire occurrence probability and area. The denominator is the total area of the district. A higher level of DFDI implies a greater propensity to burn. In evaluating the effectiveness of alternative spatial strategies, a preferred proposal is one that reduces the DFDI more.

The second measure, Moran's I coefficient of spatial autocorrelation (Cliff and Ord 1981), evaluates the neighborhood effects of fires. Formally, the I coefficient is defined as:

$$I = \frac{n}{2a} \; \frac{\Sigma_i \Sigma_j w_{ij} (x_i - \bar{x})(x_j - x)}{\Sigma_i (x_i - \bar{x})^2} \tag{4}$$

where n is the number of geographic units; $w_{ij}$ is the contiguity weight of spatial relationship, which equals one if the i-th and j-th geographic units are adjacent to each other and zero otherwise; $X_i$ is the study variable, which is the probability of fire occurrence in this study; 2a is the sum of contiguity weights of all pairs of geographic units, i.e., $2a = \Sigma_i \Sigma_j w_{ij}$.

In general, a positive value of the I coefficient indicates a pattern of positive spatial autocorrelation where areas of similar values tend to be close to one another. A negative I coefficient indicates a pattern of negative spatial autocorrelation where areas of high values tend to be located near areas of low values. When the I coefficient is close to zero, the pattern is considered random.

In terms of fire danger, a clustered pattern of fire occurrence probability is least desirable because it implies a distribution where areas of high burn-propensity are concentrated thereby maximizing the threat of wildfire holocausts; a scattered spatial pattern is preferred because it separates areas of high burn-propensity into isolated patches. In a pattern of negative spatial

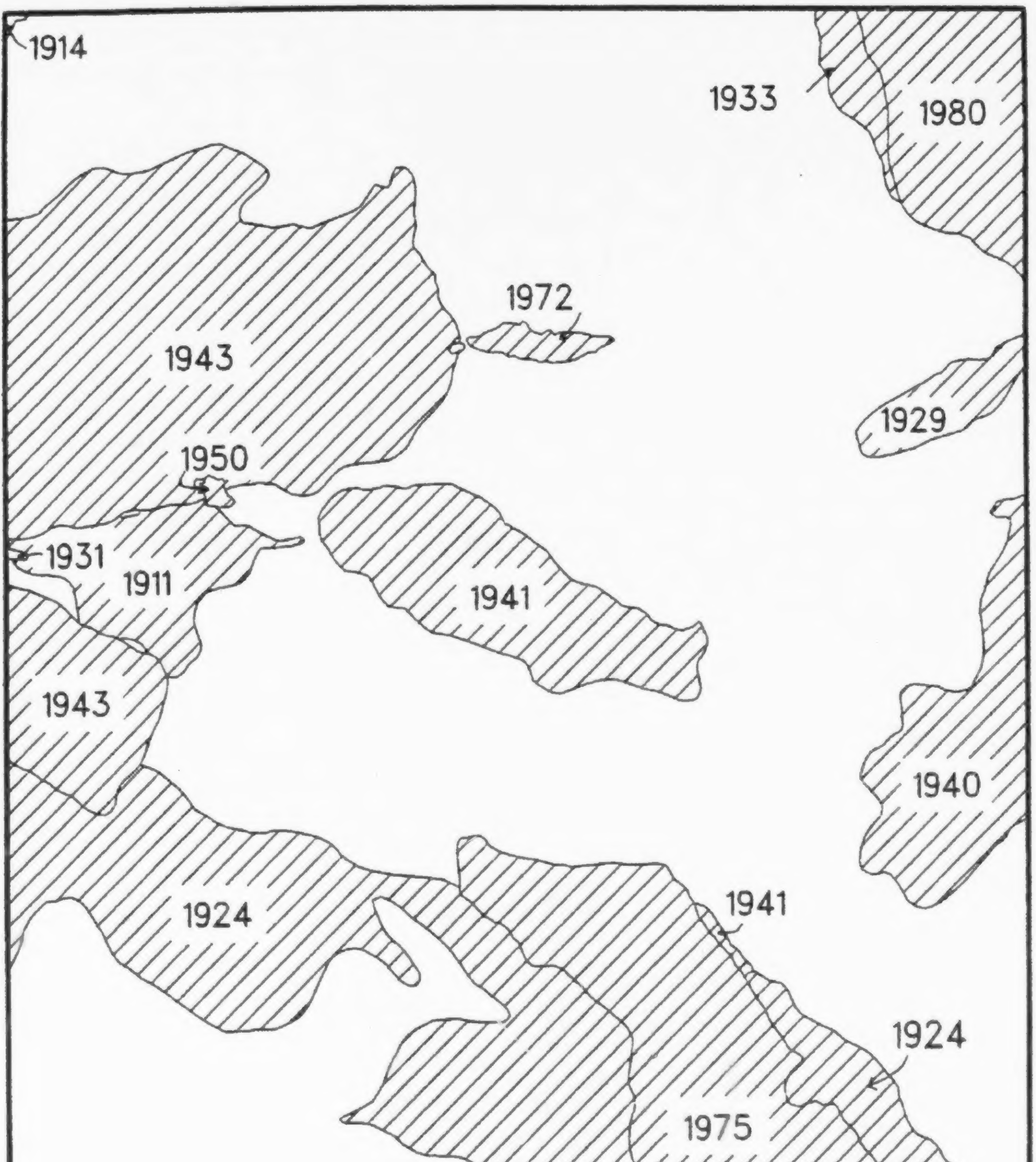

1914
1933
1980
1972
1943
1929
1950
1931
1911
1941
1943
1940
1924
1941
1924
1975

Figure 2. Fire activity between 1911 and 1984.

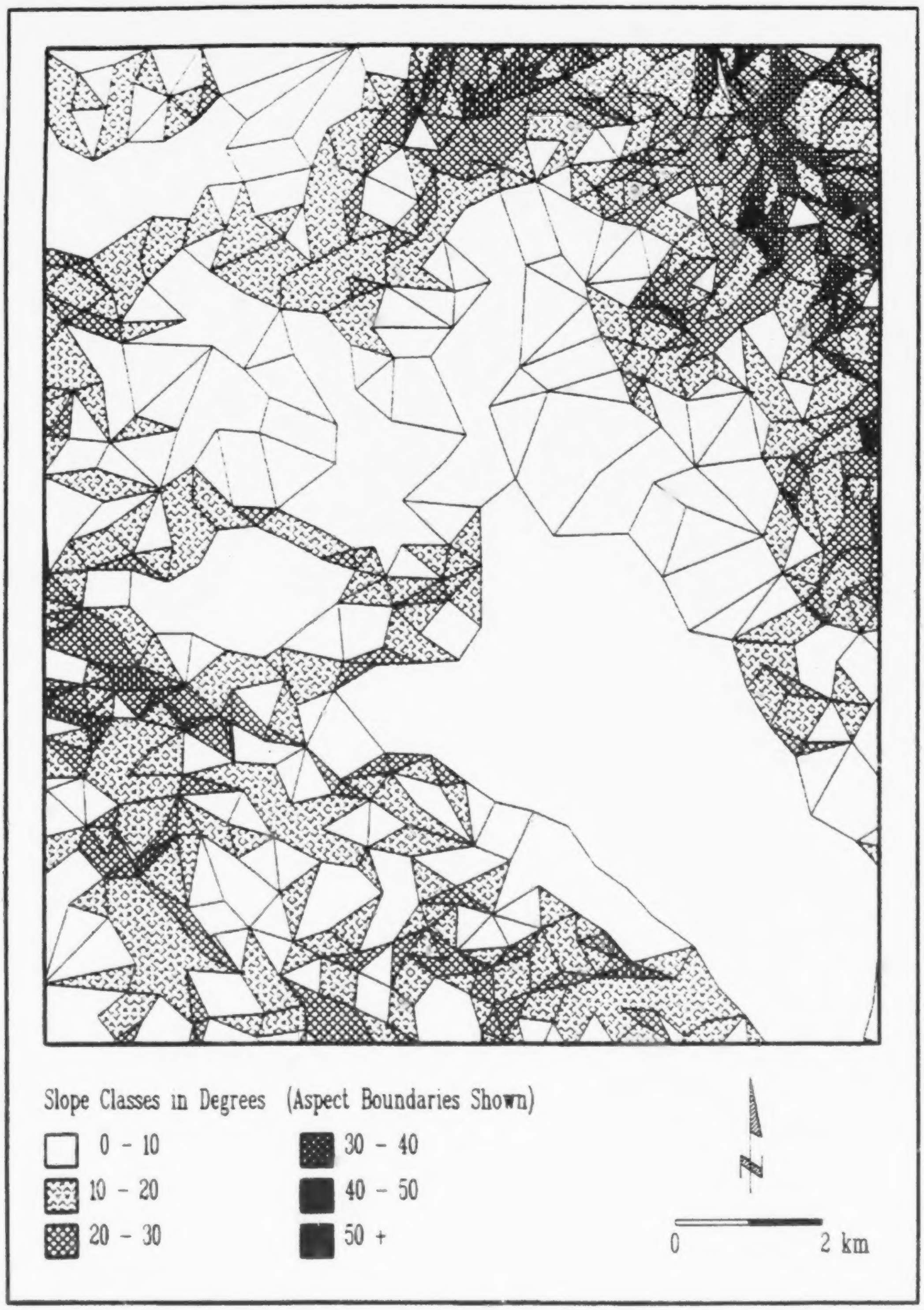

Slope Classes in Degrees (Aspect Boundaries Shown)
0 - 10
10 - 20
20 - 30
30 - 40
40 - 50
50 +
N
0
2 km

Figure 3. Geographic units defined by topographic characteristics.

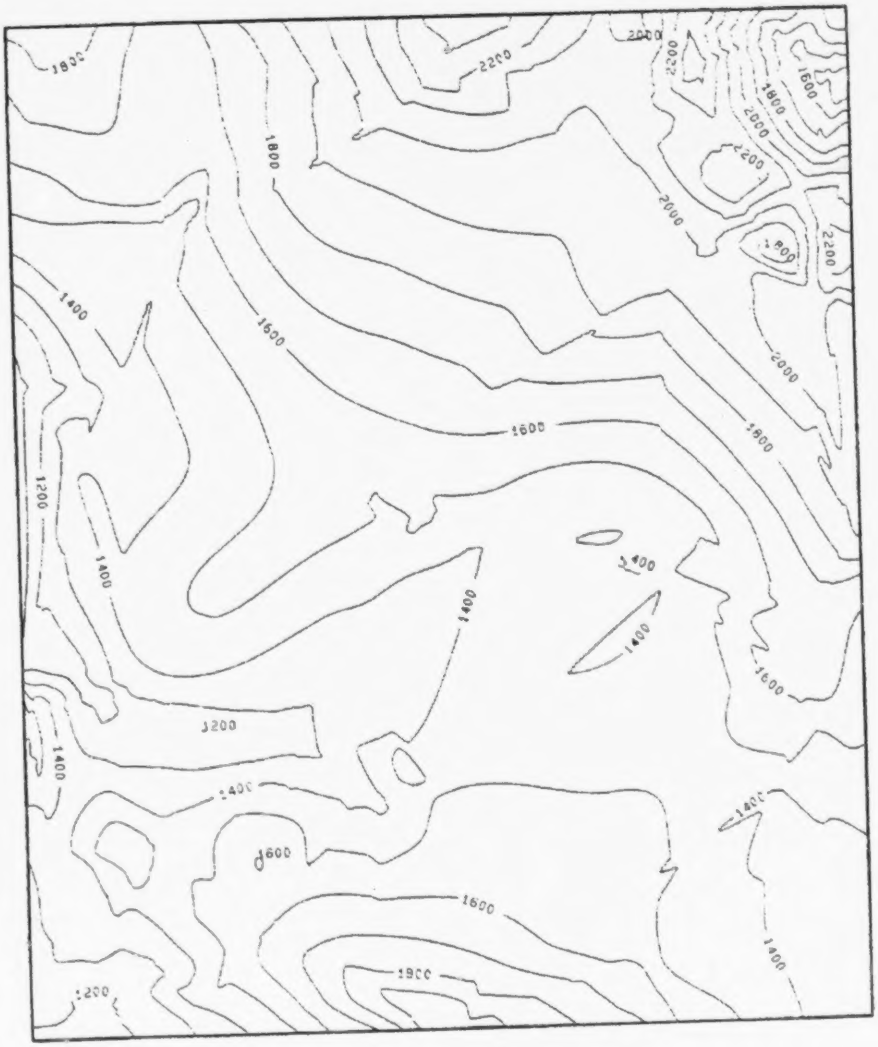

1800
2200
2000
2200
1600
1800
2000
2200
1800
2000
2200
1800
1400
1600
2000
1600
1800
1200
1400
1400
1400
1600
1200
1400
1400
1400
1600
1600
1400
1800
1200

Figure 4. Topography of the study area.

autocorrelation, each parcel of high fire danger is surrounded by parcels of low burn-propensity, thus the chance for a wide spread, uncontrollable fire is minimized. In evaluating spatial strategies of prescribed burning, the strategy resulting in a lower value of the I coefficient is preferred.

The cost of a strategy consists of two factors: total area to burn and proximity to roads. In general, a larger area costs more to burn than a smaller area. Also, burning areas that are closer to roads is cheaper because it is easier to transport crew and equipment.

## Evaluating spatial strategies

The effectiveness of a spatial strategy was evaluated according to both DFDI and Moran's I. The DFDI was calculated for each of the eight strategies listed in Table 1. Based on $DFDI_0$, the DFDI of the original map, RDFDI is defined as the degree to which a strategy can reduce DFDI from the pre-burned state. Therefore, for the strategy k,

$$RDFDI_k = (DFDI_0 - DFDI_k)/DFDI_0$$

Also, EDFDI is defined as the effect of the strategy in reducing DFDI, relative to the strategy of maximum RDFDI among m alternative strategies, such that

$$EDFDI_k = RDFDI_k / MAX\ RDFDI_m$$

Likewise, Moran's I coefficient was calculated for each strategy and defined EI as the effect of a strategy in producing a more desirable (scattered) pattern, relative to the minimum of this measure, such that

$$EI_k = 1 - (I_k - MIN\ I_m)/MIN\ I_m$$

Finally, the effect of a specific strategy, E, is defined as the product of EDFDI and EI such that

$$E_k = EDFDI_k \ X\ EI_k$$

The calculated E indices for the eight strategies are listed in Table 1. According to this measure, the sixth strategy which will burn polygons 763, 774, 803, and 844 is most effective in reducing the overall fire danger of the district and producing a more desirable spatial pattern.

The cost of a spatial strategy is determined by the total area to burn and proximity to roads. It is well-known that the relationship between area and cost is nonlinear, i.e., it costs less than twice as much to burn an area twice as large. For this reason, the cost in terms of area, CAREA, is defined as the square root of the area to burn, relative to the maximum of this measure, such that

$$CAREA_k = (A_k /MAX\ A_m)^{1/2}$$

Likewise, the cost in terms of proximity to roads, CROAD, is defined as the square root of the average distance to a nearest road, relative to the maximum of this measure, such that

$$CROAD_k = (ROAD_k /MAX\ ROAD_m)^{1/2}$$

Finally, the cost factor for each strategy, C, is derived from the product of CAREA and CROAD such that

$$C_k = CAREA_k \ X\ CROAD_m$$

The calculated C indices for the eight strategies are listed in Table 1. According to this measure, the second strategy which will burn polygons 257, 344, and 404 is most costly and the eighth strategy which will burn polygons 919, 948, 955, and 977 is least costly.

To take into account both effectiveness and cost in evaluating spatial strategies, the measure of cost-effectiveness, E/C, is defined as the ratio of E to C. The calculated E/C indices for the eight strategies are listed in Table 1.

Based on the E/C measure, the fifth strategy is most cost-effective while the second strategy is least cost-effective. Accordingly, the highest priority of prescribed burning should be assigned to the fifth strategy.

## Conclusion

To effectively manage wildland fires in a large district under a limited budget, one must evaluate the cost-effectiveness of all possible management plans and set the priority of alternative spatial strategies. It is important to realize that not all prescribed burns are desirable, and that certain alternatives could be less expensive yet more effective. Although some prescribed burnsmay be important locally, if they are not located and scheduled according to a well-designed, full-scale district management plan, the expensive efforts of prescribed burning may simply create a situation of potential fire disaster in the future.

The recent development in GIS technology has made it possible for wildfire managers to evaluate possible spatial strategy of prescribed burning. Using a GIS, we can establish an ecological database which incorporates data of fire activity, type and age of vegetation, topographic characteristics, weather conditions, human structures, and existing preventive treatments. From the database, a logistic model of fire occurrence probability can be constructed. The distribution of fire occurrence probability can be mapped by calibrating the best-fit model of fire occurrence, which then allows for the identification of critical zones of fire danger. Different spatial strategies can be evaluated effectively and a comprehensive management plan can be obtained.

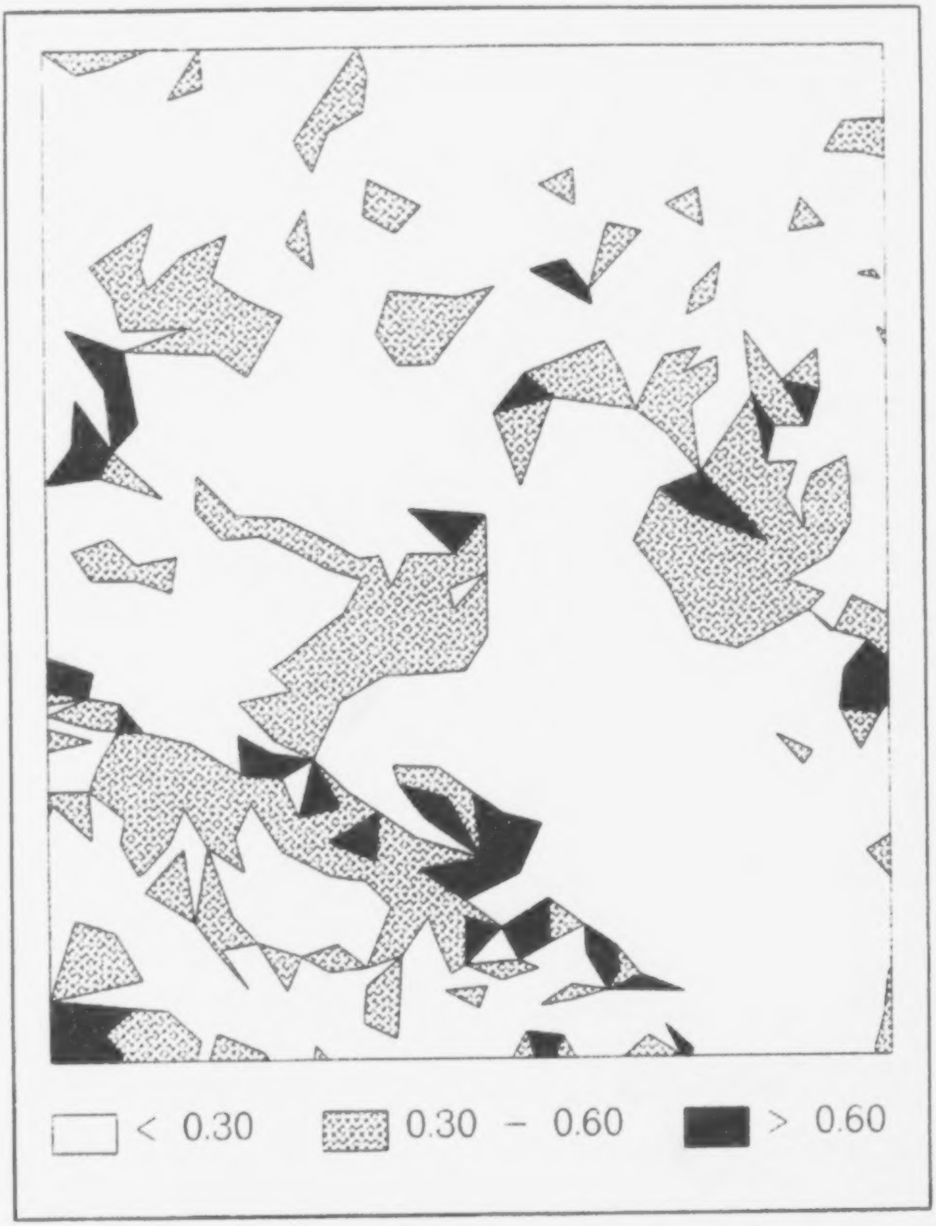

< 0.30
0.30 – 0.60
> 0.60

Figure 5. Classification of geographic units by fire danger.

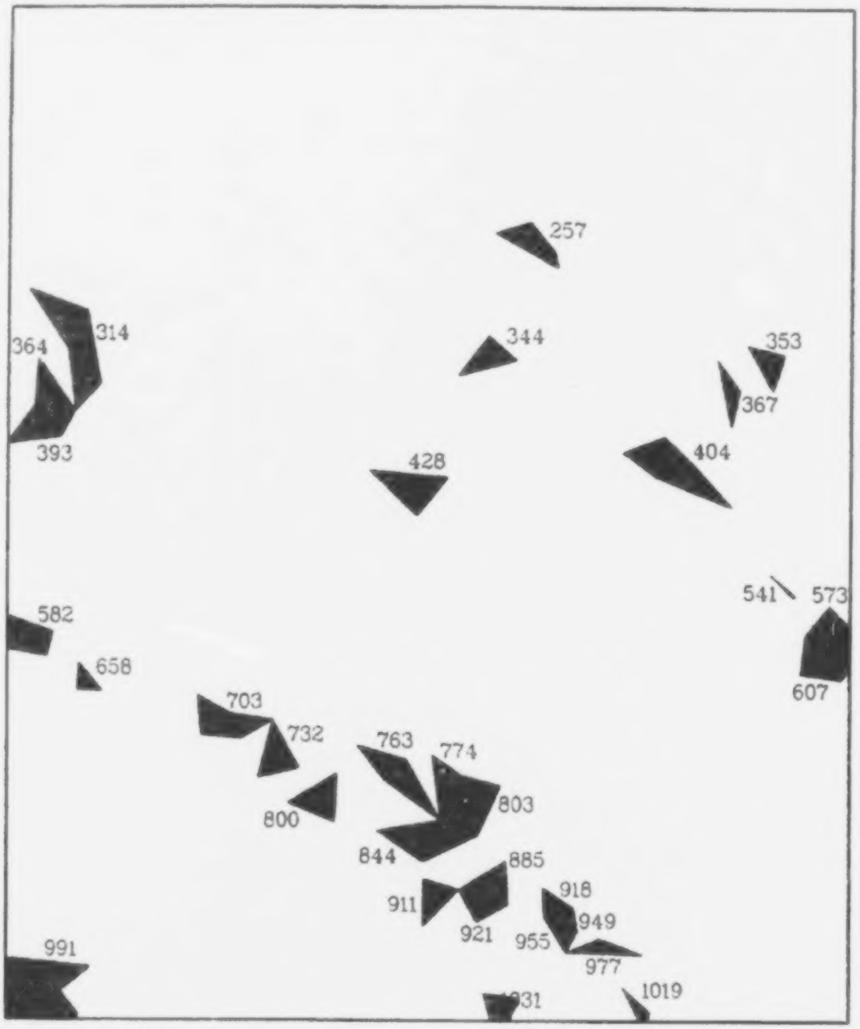

257
364
314
344
353
367
393
428
404
541
573
582
607
658
703
732
763
774
800
803
844
885
911
918
949
921
955
977
991
31
1019

Figure 6. Eight spatial strategies of prescribed burning.

## Acknowledgment

This study is sponsored by the U.S. Forest Service under grant PSW-91-0042CA. I am grateful to R.A. Chase of Riverside Fire Research Laboratory for his support of this study.

## References

BMDP P-series (1987) P. 330-343. Program PLR. University of California Press.

Chou, Y.H. (1992) "Wildfire management with a GIS." *International Journal of Geographical Information Systems*, (in press).

Chou, Y.H., R.A. Minnich, L.A. Salazar, J.D. Power, and R.J. Dezzani (1990) "Spatial autocorrelation of wildfire distribution in the Idyllwild quadrangle, San Jacinto Mountain, California. *Photogrammetric Engineering and Remote Sensing*. 56: 1507-1513.

Cliff, A.D. and J.K. Ord (1981) *Spatial Processes: Models and Applications*, Pion.

Emrick, W.E and T.E. Adams (1977) *Brush Management - Use of Prescribed Fire*, Cooperative Extension, U.S. Department of Agriculture, University of California, Berkeley, California.

Environmental Systems Research Institute (1987) ARC/INFO TIN User Manual. ch. 2, Fundamentals of TIN. Redlands, California.

Green, L.R. (1981) *Burning by Prescription in Chaparral*, General Technical Report PSW-51, Pacific Southwest Forest and Range Experiment Station, U.S. Forest Service, Berkeley, California.

U.S. Geological Survey (1987) Digital Elevation Models. USGS DataUser's Guide. 5: 38.

Wrigley, N. (1976) "An introduction to the use of logit models in geography." *Concepts and Techniques in Modern Geography*, 10. Norwich: Geo Abstract Ltd.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# The application of GIS technology for forest fire management planning

**Bryan S. Lee**

Forestry Canada
Northern Forestry Centre
5320 - 122 Street
Edmonton, AB, Canada T6H 3S5

**David J. Buckley**

TelTec, GIS Division
Suite 210 - 1430 Florida Avenue
Longmont, CO, USA 80501

**Mark Kressin**

IBM Corporation
Federal Sector Division
6300 Diagonal Highway
Boulder, CO, USA 80302

## Abstract

*In recent years Canadian forest fire management agencies have become more dependent on decision support systems for planning and real time decision-making. In particular, the integration of GIS technology and expert systems has become increasingly important for realistic modelling of forest fire potential and fire behaviour situations. However, to date few operational systems have been successfully implemented in an effective and cost-efficient manner. The Intelligent Fire Management Information System (IFMIS), developed by Forestry Canada, is one exception.*

*Following on the design and principles of IFMIS, a prototype was developed using several different GIS software offerings to evaluate the feasibility and issues related to implementing a fully operational GIS forest fire preparedness planning tool. Specific models were developed using typical GIS toolkit functions and application programming interfaces (API) to address basic fire management concerns. These applications utilize weather, forest inventory, fuels, base map, and topographic data in a real time integrative fashion. A range of different application scenarios were developed to address fire preparedness planning requirements including fire weather modelling, fuel table modelling, fire behaviour prediction, fire detection and the deployment of initial attack resources. Both the Canadian Forest Fire Weather Index System (FWI) and the Canadian Forest Fire Behaviour Prediction System (FBP) were incorporated as the platform for all empirical modelling equations.*

*This paper presents the results of the prototype, including a review of the issues related to integrating GIS technology and expert systems techniques in a manner consistent with fire management methodologies and requirements. An emphasis is placed on the opportunities for customizing commercial GIS software using industry-standard toolkits. Of particular concern is the issue of temporal data. This involves both historical fire data and time-sequenced fire growth predictions. Methodologies will be described that afford both regular monitoring of fire potential as well as real time prediction of fire behaviour. Examples will be presented including a real time demonstration if time permits.*

# B5 Harvesting operations

**Integration of silvicultural record keeping and geographic information systems**

Robert Kennett, Reid, Collins and Associates, Vancouver, BC, Canada

**GIS-based strategic planning at MacMillan Bloedel Limited**

Brad Whitehead et al., MacMillan Bloedel Limited, Nanaimo, BC, Canada

**Elements of management for tracking forest planning in Quebec with the aid of GIS**

Gilles Arseneau et al., Université Laval, Sainte Foy, PQ, Canada

**Probability based spatial-temporal modelling for natural forests: sampling techniques, model development, accuracy assessment, and residual evaluation**

Kim Lowell, Université Laval, Sainte Foy, PQ, Canada

GIS'92

Symposium

Vancouver
British Columbia
February 1992

# Integration of silvicultural record-keeping and geographic information systems

**Robert Kennett, RPF**

Reid, Collins and Associates
A Division of H.A. Simons Ltd.
503-425 Carrall Street
Vancouver, B.C.
V6B 2J6

## Abstract

*Conceptually, silvicultural record-keeping is an ideal application for Geographic Information Systems. Silvicultural activities are spatial in nature, are linked to other thematic layers such as forest inventory, and contain descriptive attribute data. However, implementation of a silvicultural record-keeping system within the framework of a GIS is not a straight forward task due to the nature of the data.*

*This paper examines the motivation for silvicultural record-keeping and outlines some of the problems that may be encountered when integrating silvicultural information into a GIS. A conceptual framework for development of a record-keeping system is also presented.*

## Introduction

Geographic Information Systems have been used to maintain forest inventory information for a number of years. More recently, GIS has been used in other forestry applications. One application that is attracting greater attention is silvicultural record-keeping systems.

Traditionally, silvicultural records have been maintained either in a manual filing system, or in a non spatial computerized database system. However, there are three characteristics of silvicultural records that make them ideally suited to implementation within a GIS framework.

First, silvicultural activities are spatial in nature. They take place on polygonal areas of land that are defined in the real world with areas and locations. Second, each activity has descriptive attributes (e.g., species, number of trees planted, etc.) associated with it.

Finally, silvicultural information is linked with other thematic layers in the database. For example, forest inventory information (species, stocking, quality) is dependent to some degree on the silvicultural activites that take place. Also, the actual treatments that are performed are a function of other parameters such as slope, aspect, elevation, biogeoclimatic zone, and susceptibility to insects and disease.

While the concept of a GIS-based silviculture record-keeping system is not complicated, its actual implementation can be extremely complex due to the regulatory environment and the variety of silvicultural treatments that can be applied to a single area.

This paper examines the motivation for developing a GIS-based silvicultural record-keeping system, outlines the problems associated with such a system, and suggests a framework that may be used.

## Motivation for silvicultural record-keeping system

There are several reasons why forest companies are becoming more interested in implementing a GIS-based silvicultural record-keeping system. The most important of these reasons is the ever increasing reporting requirements of the provincial government. These requirements result from the method by which forest land is administered in British Columbia. Ownership of the majority of forest land is retained by the government, with forest companies harvesting the timber from this land under a variety of administrative tenures. These tenures usually require reforestation to the "free-to-grow" state following harvest. There may also be provisions in the tenures for incentives to carry out more intensive silvicultural activities such as thinning and fertilization.

As a result, the most common question asked by our clients when discussing record-keeping is "Can it do MLSIS reports?" MLSIS, or *Major Licencee Silvicultural Information System,* is the mechanism used to report silvicultural activities on Crown lands in British Columbia. This system consists of standardized reports and maps that are submitted three times per year for harvesting operations, and four times per year for silvicultural operations. These reports and maps require maintenance of accurate information concerning silvicultural activities, and usually involve a staff member going through the files for several days to complete the requirements.

In addition to the MLSIS reports, each licencee must submit an annual report to the Ministry of Forests describing the operations (including silviculture) that have taken place during the previous year. A computerized record-keeping system is a very valuable tool for preparing this report.

Along with the requirement for reporting to government agencies, there are also other external as well as internal reporting requirements that must be satisfied. One of our clients reports that questionnaires relating to silvicultural costs and treatment areas were received from the Canadian Pulp and Paper Association, Council of Forest Industries, and Ministry of Forests last year. Internally, managers often need to know the questions such as how many hectares were planted last year, and how much did it cost? These questions may be required because of traditional accounting and budgeting processes, or because the public and special interest groups are starting to raise these types of questions more frequently. Usually, the managers need the answers "yesterday" which implies that the information must be readily at hand.

The third reason that forest companies implement silvicultural record-keeping systems is that they can assist in the planning of work programs for the future. For example, foresters need to know how many hectares of land will be planted next year, and where the planting will be located so they can determine budgets, organize the planting contracts, and check on accessibility.

Another reason for implementing a record-keeping system is to provide better forest management through the process of learning from past successes and failures. At times, it has seemed that success was measured by the number of wheelbarrows it took to carry their completed field survey forms. A wealth of information was collected but it was difficult to use because it couldn't be easily accessed. As professionals, foresters are constantly trying to improve their knowledge and performance. If accurate, easily accessible records are maintained it is possible to answer questions like "What combination of treatments and physical factors make certain plantations perform above average? Can they be duplicated elsewhere?"

The dollars spent on silviculture by governments and companies represent a very significant investment. One company has used the following analogy to emphasize the need for GIS-based silvicultural record-keeping systems:

> *"If a company were to take the money it spent on silviculture and invested in the stock market, the shareholders would want to know the status of that investment. Why would we treat our investment in silviculture any different from an investment in the stock market?"*

Finally, the different working environment of today provides a motivation for maintaining computerized record-keeping system. The work force is much more mobile now than it used to be, with the result that there are frequent staff turnovers. It is no longer possible or acceptable to allow staff members to retain the knowledge in their heads or in a filing system that only they understand.

## Requirements of a GIS-based silvicultural record-keeping system

Several fundamental concepts must be emphasized in order that a GIS-based silvicultural record-keeping system can be successfully implemented in a real world situation.

The system must be developed with an open design that will allow the system to be easily modified as new types of treatments are incorporated into the silvicultural prescription and the reporting regulations change. Changes in reporting regulations may require only a change in report formats; however in some cases they may require new data to be collected and thus stored in the database.

One of the advantages of computerized database systems is the amount of information that can be stored. Unfortunately, it is not often realized by the user that there is a cost in terms of system complexity and management that is associated with these great volumes of data. Decisions regarding which data and at what level of detail is to be collected must be made in advance.

An often heard response to this is "Well, you never know ...someone might want to know that" which usually results in a system full of data that is never accessed. A good rule to follow is that if you cannot identify and document a particular need then the data should not be in the system. If a future need for this data arises, there should not be a difficulty in incorporating it into the system if an open and flexible design is used.

## Problems associated with implementing a GIS-based silvicultural record-keeping system

As discussed in the introduction, maintenance of silvicultural activities is an application that appears to be ideally suited to implementation within the framework of a GIS. However, there are a number of characteristics of silvicultural activities that make successful implementation of a GIS-based system difficult.

First, there are many different types of activities that can take place within a forest company's operating area. For example, typical silvicultural activities might include site preparation, regeneration, brushing and weeding, regeneration surveys, free-to-grow surveys, juvenile spacing, pre-commercial thinning, commercial thinning, pruning, and fertilization. Each one of these activities has a different set of descriptive attributes that must be maintained. For example, site preparation attributes might include the method, cost, and degree of disturbance created. On the other hand, regeneration attributes may include species, type of stock, seedlot, planting density, seedling quality, and cost. For this reason, it is impossible to store the attributes for all activities in a common database table, which requires that a thematic layer must be created for each different activity type.

Further complicating the problem described above is the fact that it is necessary to examine the different types of activities together. Foresters often ask questions such as "Which areas of land were planted in 1985 and have had free-to-grow surveys completed?". Therefore, it is not sufficient to create individual thematic layers for each of the different activity types. Rather, a resultant layer composed of the overlay of each of the different activity types must be created.

This requirement for a resultant thematic layer presents the second major problem in implementing a GIS-based silvicultural record-keeping system. Different silvicultural activities are seldom carried out on exactly the same area of land, with the consequence that many valid resultant polygons, and even more invalid "sliver" polygons are created. Figure 1 shows how these sliver polygons can be created.

The valid resultant polygons present problems in ensuring that the correct attributes are stored with each one. Although most GIS systems supposedly handle this aspect automatically, very few actually make it easy to implement. Because silvicultural records are dynamic in nature, it is necessary to recreate the overlays each time a new activity takes place.

The invalid sliver polygons present an even more complicated question. When is a sliver polygon not a sliver polygon? Is it 0.5 hectare, 0.25 hectare, or 1.0 hectare? An appropriate answer is that the data should only be kept down to the level at which the resource is managed. The difficulty is that there is a reluctance to "throw out" this "data", even though it will never be used at an operational level. Also, the elimination of the slivers must be consistent so that the same database is created if the overlay process is repeated.

The third problem associated with implementing a record-keeping system is that the same type of activity can occur to the same area of land more than once. For example, plantations often fail and it is necessary to replant all or part of the area originally planted. If the regeneration activity type is treated as a single thematic layer, it becomes very difficult for a system to distinguish and keep track of the two individual activities, particularly if the two activities are not entirely coincident.

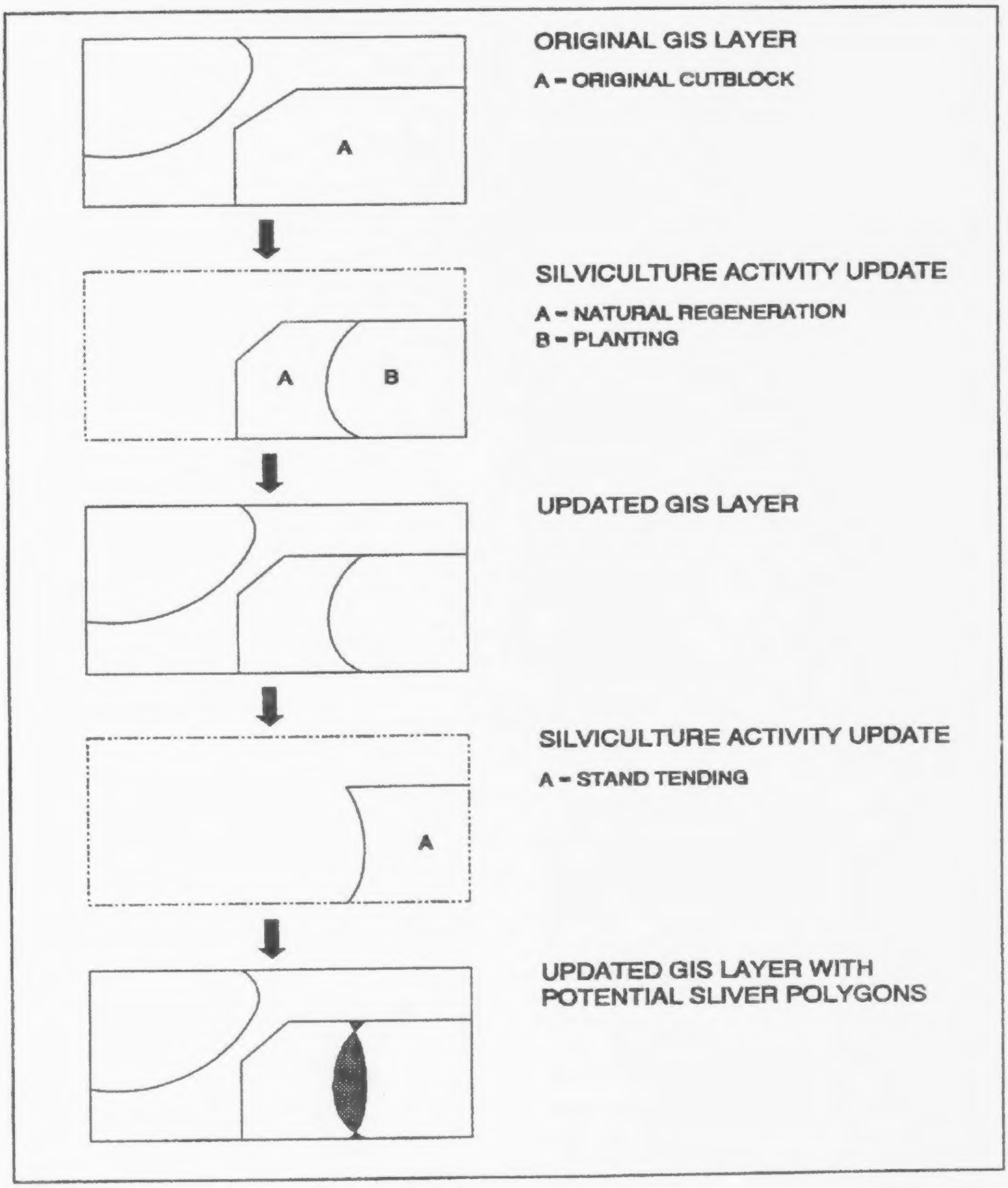
A fourth problem of integrating a silvicultural record-keeping system into a GIS database is the procedure by which the forest inventory thematic layer is updated. When a cutover area becomes "free-to-grow", the inventory database is updated. However, inventory polygons are based on homogeneity of factors such as species, age, and height. Areas within a cutover area may have received totally different treatments to achieve the "free-to-grow" status, but may have identical stands from an inventory perspective. Thus, the process of converting from silvicultural records to inventory must be capable of undertaking the generalization necessary to create these homogenous polygons, while also maintaining a record of the individual treatments that created them.

This last problem raises another issue. To what degree does a silvicultural record-keeping system also become a history record-keeping system? If part of its purpose is to maintain a record of the history of a given area of land, should it also maintain a history of previous forest types and treatments? This may not be a serious problem where clearcutting and long rotation ages are prevalent. However in situations where partial cutting is used or in tropical plantation forestry where effective rotation ages can be as short as 5 years, it will be a significant issue in terms of data management and data volumes.

A final problem, at least for developers of a GIS-based system, is the need to create a solution that will work for more than one client. In this way, development costs can more readily be recovered, and the purchase price for the end-user reduced. Unfortunately, there is no standardization of GIS software or database software between users. Furthermore, the reporting regulations are different in each jurisdiction and the types of activities, as well as the individual attributes recorded for each activity always seem to be unique for each user.

**FIGURE 1**

**CREATION OF SLIVER POLYGONS THROUGH GIS UPDATE PROCEDURES.**


ORIGINAL GIS LAYER
A – ORIGINAL CUTBLOCK
A
SILVICULTURE ACTIVITY UPDATE
A – NATURAL REGENERATION
B – PLANTING
A
B
UPDATED GIS LAYER
SILVICULTURE ACTIVITY UPDATE
A – STAND TENDING
A
UPDATED GIS LAYER WITH POTENTIAL SLIVER POLYGONS

## Sugested framework for implementing a GIS-based silvicultural record-keeping system

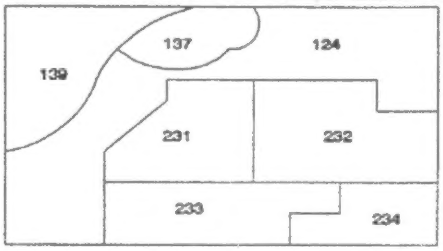
The intent of the framework proposed below is to promote discussion and exchange of ideas for successful implementation of a GIS-based record-keeping system. It is not intended to present the best solution to the problems described above, or even to solve all of these problems.

The basic administrative unit used in the system is the cutblock, or opening. Each cutblock entered into the system has a unique identifier, and all silvicultural activities are referenced to a specific cutblock. Each activity that takes place within the cutblock must be assigned a unique number or other identifier in order to facilitate activities of the same type occurring in the same opening.

The attribute data is stored in a series of related database tables; "master table", "base data table", and individual "activity data tables". Figure 2 shows the three types of tables and linkage to the graphic database. In this example, there are seven polygons, three of which (124, 137 and 139) contain standing timber and have had no silviculture treatments; thus they have no reference in the silviculture database. The remaining four polygons are formed as the result of harvesting and silvicultural activities. Polygon 231 is formed from a single opening (A112), while polygons 232, 233 and 234 are formed from a second opening (A143) and the subsequent silvicultural activities.

FIGURE 2

SILVICULTURE ATTRIBUTE TABLES.



MASTER TABLE

| POLYGON# | OPENING# | ACTIVITY1# | ACTIVITY2# | ACTIVITY3# . . . | AREA |
|---|---|---|---|---|---|
| 231 | A112 | R012 | S005 | .... | 10 |
| 232 | A143 | R014 | | | 14 |
| 233 | A143 | R014 | S003 | .... | 11 |
| 234 | A143 | S003 | | | 8 |

BASE DATA TABLE

| OPENING# | ATTRIBUTES | | | |
|---|---|---|---|---|
| A112 | NW | 375 | 10 ha | .... |
| A143 | N | 350 | 33 ha | .... |
| ⋮ | ⋮ | ⋮ | | |

REGENERATION TABLE

| ACTIVITY# | ATTRIBUTES | | | |
|---|---|---|---|---|
| R012 | FIR | 1978 | 10 ha | .... |
| R014 | PINE | 1979 | 25 ha | .... |
| ⋮ | ⋮ | ⋮ | | |

STAND TENDING TABLE

| ACTIVITY# | ATTRIBUTES | | | |
|---|---|---|---|---|
| S003 | HAND | 1980 | 19 ha | .... |
| S005 | SPRAY | 1979 | 10 ha | .... |
| ⋮ | ⋮ | ⋮ | | |

The master table forms the linkage between the spatial database and the attribute tables. There is one record in this table for each of the resultant polygons formed during the overlay of the various silvicultural activity themes in the spatial database. The actual attribute data that is stored in this table consists of:

- area
- pointers to the activities that have taken place on the polygon in question

The base data table contains attribute information that applies generally to the entire cutblock, and is relatively static over time. Therefore, only one record per cutblock is required. Typical types of attributes that are stored in this table include:

- total cutblock area
- slope
- aspect
- elevation
- biogeoclimatic zone
- administration (forest region, district, tenure, etc.)

The activity data tables contain the descriptive attributes for each activity. The different specific silvicultural activities can be grouped into general categories of treatment types for which there are similar attributes that must be stored. For example, "REGENERATION" is a general activity type, and specific activities that occur within this general type may be planting, natural regeneration, seeding, fill-in planting, and coppicing. Each of these specific activities contains similar attributes such as species, stocking, etc., and can therefore be grouped together into one attribute table called "REGENERATION".

Through this process of grouping into general activity types, it is possible to reduce the number of different activity attribute tables required to approximately six, including:

- denudation
- site preparation
- regeneration
- surveys
- brushing/weeding
- stand tending

There is one record per activity within the cutblock, and that record is entered in the appropriate general activity table in the list above. The specific attributes that are entered for each activity will vary with the general activity type, but there are some fields that are common to all types.

Some examples of these include:

- specific activity name (e.g., planting, natural regeneration)
- activity identification number
- date
- actual or planned
- area (from field traverse)
- cost

There are a few things to note with regard to the actual attributes stored for each activity. First, the activity can either be planned for some point in the future, or have already happened. In this way, reports can be generated that will outline proposed future silvicultural programs. The process of assigning future activities may be automated to a certain extent (e.g., surveys of some description usually follow regeneration), but some manual assignment and checking by a forester is also required.

Second, the area for the activity is entered by the user based on a field traverse or some other measure, rather than being calculated by the GIS system. The reason for this is that unless a more precise method than digitizing (e.g. coordinate geometry) is used to enter the boundaries of the activity, the area calculated by the system will not be the same as that calculated when the activity takes place due to the inherent error involved in the transfer and table digitizing operations. Furthermore, there will likely be areas that are significant from the regulatory point-of-view and which add nothing to the graphic information stored in the GIS. Examples of these areas might be small wet areas in a large clearcut that were not planted. Also, removal of sliver polygons will alter the area if it is calculated by the GIS. Therefore, the only way to ensure consistency on the reports submitted to the government is to use an entered area based on the field traverse. The areas calculated by the GIS would be used for any analysis or adhoc queries involving overlays of different activities.

The above paragraphs have defined the general database structure of the record-keeping system. To be integrated into the GIS environment in a manner that can be used efficiently, some additional requirements must be met.

First, the silvicultural records and the tools necessary to maintain them must be accessible within the GIS environment. At the simplest level, this means that an import/export routine should not be required to move attribute data into and out of the GIS. The database management system (DBMS) should be directly accessible from within the graphics environment of the GIS, to facilitate attribute updates and queries without moving in and out of different programs.

Equally important, both the GIS and DBMS should implement the concept of relational database structures and database views. Because the complex nature of silvicultural records requires the attribute data to be stored in several different tables, but queried as if it were one table, simultaneous access to related records in different tables is required. Most database management systems support this concept, but very few GIS packages are able to take advantage of this functionality within the graphics/query environment.

Finally, a smooth process for creation of the resultant polygons, rationalization and removal of sliver polygons, and automatic linkages between the resultant polygons and the appropriate attribute records must be implemented. Most GIS vendors will tell you that this is "no problem", but in practice developing a mechanism that does not break down when you get into a real world situation with real world data is not straightforward.

## Possibilities for the future

The framework described above is based on working within the limitations set by most GIS software packages today. However, there are significant breakthroughs being made that should allow implementation to proceed more smoothly.

Perhaps the most encouraging development is the trend toward supporting the concept of "one-to-many" relationships between the spatial and attribute databases. This process, illustrated in Figure 3, greatly simplifies the mechanism by which several different activity attribute records can be related to the same polygon in the spatial database.

Another very positive development is the trend towards transparent linkages to a variety of relational database systems by the same GIS package. This will make it far easier for developers of applications such as silvicultural record-keeping to support multiple hardware and software platforms, resulting in enhanced product quality for a reduced price.

Also very encouraging is the trend by GIS vendors to take advantage of the advanced features built into most DBMS systems, including relational structures, views, and input screens. This tighter integration of the DBMS into the GIS environment can only result in more user friendly applications that make it far easier to maintain a valid, consistent database.

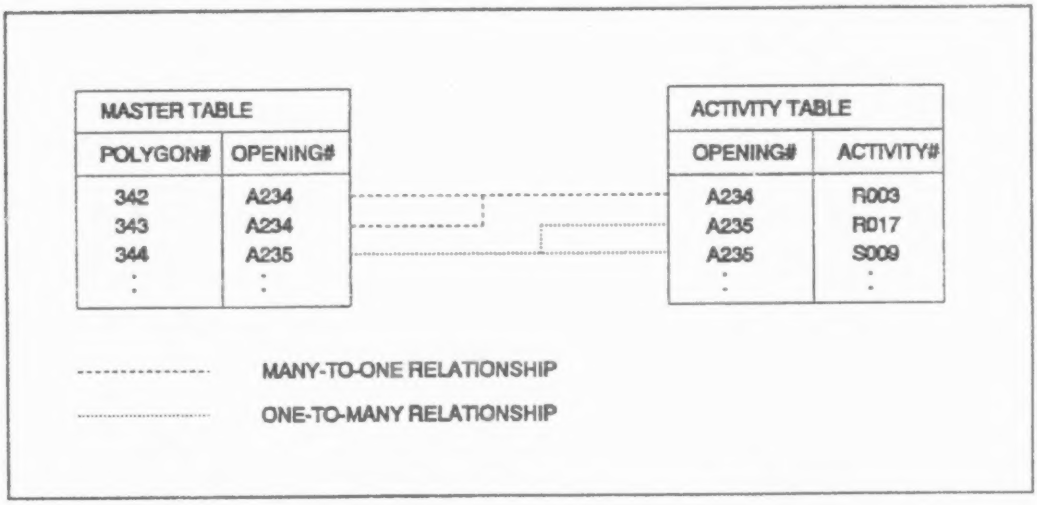
## Conclusions

Although the implementation of a GIS-base silviculture record-keeping system is quite complex, many of the difficulties that have been outlined in this paper can be overcome through careful planning and through close cooperation between the users and the developer. Above all the developer must understand the user's needs and the users must also have a thorough comprehension of their own requirements and priorities.

The current focus of GIS development and integration with database management system capabilities is extremely positive as it can only lead to easier and less restrictive application development.

FIGURE 3

MANY-TO-ONE AND ONE-TO-MANY FILE RELATIONS



MASTER TABLE

| POLYGON# | OPENING# |
|---|---|
| 342 | A234 |
| 343 | A234 |
| 344 | A235 |
| : | : |

ACTIVITY TABLE

| OPENING# | ACTIVITY# |
|---|---|
| A234 | R003 |
| A235 | R017 |
| A235 | S009 |
| : | : |

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# GIS-based strategic planning at MacMillan Bloedel Limited

**Peter J. Kofoed**

MacMillan Bloedel Limited
Woodlands Services Division
Nanaimo, BC V9R 5H9

**Glen A. Jordan**

Faculty of Forestry
University of New Brunswick
Fredericton, NB E3B 6C2

**Brad T. Whitehead**

MacMillan Bloedel Limited
Woodlands Services Division
Nanaimo, BC V9R 5H9

### Abstract

*MacMillan Bloedel Limited, like many Canadian forestry companies, is no stranger to long-range strategic forest planning, nor GIS technology. However, it has only recently begun to apply GIS in strategic planning. This paper documents the company's initiatives and experiences to date. It begins with a review of the company's simulation modelling approach to wood supply analysis (forest estate modelling) during the 1980s. The review pays particular attention to difficulties associated with lack of location-specific resolution. The paper follows with a discussion of two recent Company initiatives in applying GIS in forest modelling. In the first, a grid-based approach is introduced and offered as a better alternative to vector-based approaches. In the second, a start has been made on incorporating spatial constraints, including adjacency conditions, in the modelling process. The paper concludes with an indication of future plans that includes combining features of the two initiatives.*

*Key words: GIS, forest modelling, wood supply, strategic planning.*

## Introduction

*A forest management challenge:*
MacMillan Bloedel (MB) faces two forest management challenges common to the British Columbia Coastal forest industry. First, forests are increasingly being managed for a wide range of values. How can strategic plans be developed that are consistent with operationally integrated resource management? How may the trade-offs between options be objectively examined?

Second, the coastal industry faces the transition from virgin forest areas known as old growth to a different forest on previously cutover areas known as second growth. Old-growth timber is large, high valued, changes slowly and occurs in large undeveloped blocks. Second-growth timber is smaller, lower valued, grows rapidly and occurs in a pattern of developed stands. For MacMillan Bloedel operations, the timing of the transition varies from now through to several decades in the future depending on logging history. The challenge is to plan the transition recognizing differences (including spatial elements) between the two forests.

To meet these challenges requires recognition of the location of different forest conditions and their relative positions. This paper describes MB's recent action and outlook on GIS-based forest planning.

The first section briefly describes current forest level modelling at MB, including recent developments. It is noted that spatial identity is lacking. Discussion details the need for location-specific resolution in the planning process and then describes the foundation for such analysis—MB's GIS-based forest inventory. In the following section, two recent initiatives in spatial analysis are discussed. In the first, which began as an introduction for MB in strategic forest-level spatial modelling, discussion centers on adoption of a raster data structure. The second initiative focussed on impacts of spatial constraints on harvest scheduling at the drainage level. The final section describes MB's desire to combine elements of these two initiatives and to expand their use.

## Recent strategic forest modelling experience

Since the signing of Tree Farm Licence agreements in the 1950s and 1960s, MB has been involved in strategic forest-level analysis. The demand for such analyses and their complexity has increased dramatically. This parallels society's increasing concerns for a wide range of forest values and the continuing transition to the second-growth forest. It also reflects the increasing amount of forest information and computing power, and development of analytical procedures. Initially, analysis was for explicit calculation of Annual Allowable Cuts (AACs). Starting in the early 1980s emphasis has turned to examining the long-term timber supply impacts of forest management scenarios, particularly landbase and silvicultural options. MB's current inventory projection simulation model named FEM has proved useful for comparing such options. A major focus has been to improve the silvicultural link between operational reality and strategic analysis. This is being achieved through application of a powerful in-house yield model (Y-XENO) and development of regeneration models based on field experience. While FEM does utilize the stand-level forest inventory data that is stored in MB's ARC/INFO GIS, it does not directly recognize the location and relative positioning of forest conditions. In larger analyses stand-level data is aggregated into strata.

Development of spatial resolution in strategic-forest planning is required to meet the planning demands of the 1990s.

Many operational harvest constraints for protecting non-timber values are location specific. These include leavestrips (greenup of cutover before harvest of adjacent blocks), special management of riparian areas depending on fish values and of forage areas adjacent to deer winter ranges and local rate-of-cut restrictions for visual and hydrologic reasons. FEM indirectly allows for these concerns by requiring additional inventory to be available for harvest. These allowances are not site specific and so may be adequate in some areas but not in others; strategically the timber may be there but operationally it may not be available for harvest. There is need for a more effective link between strategic and operational levels of planning.

Spatial information is required to integrate the different spatial resolutions of forest management and basic inventory units (i.e. stands). Management prescriptions (e.g. harvest blocks) ultimately target areas that are aggregates or divisions of forest conditions and do not follow stand boundaries necessarily. This relationship between management prescriptions and the stand inventory differs between undeveloped (not roaded) old-growth areas of larger stands and the general second-growth pattern of developed smaller stands.

Mapped schedules (e.g., wildlife areas, timber harvest, silvicultural treatments) strengthen the links between strategic and operational planning. Forest management strategy is more effectively communicated to operations through maps and feedback of operational reality is encouraged through editing of the mapped schedules (manually or by computer). Such feedback loops may be repeated as desired. This process is not available with FEM when run in strata mode (aggregations of stands) as stands contributing to managed strata can be identified only indirectly.

In summary, forest level analysis must include spatial recognition to improve "on the ground" reality in strategic analysis and to effectively communicate strategic objectives to operational planning. More specifically there are three main concerns. The first is to include the impacts of adjacency constraints in strategic analysis, in the determination of harvest levels and of forest management strategies. The second is to examine harvest blocks in a flexible manner. Harvest blocks are often not tied to stand boundaries and their size and shape may change as planning proceeds. Procedures are required for grouping and partitioning stands into harvest blocks where they are not explicitly defined. The third concern is to provide mapped harvest and treatment schedules, with the ultimate objective of allowing for interactive editing of schedule maps displayed on a computer monitor.

A GIS-based inventory is of course the key to improving the exchange of the right information between strategic and operational planners. MB first started its GIS-based forest inventory in 1979, but like many companies who began early in the game it has been a long tough road and only in the last few years has the inventory been readily available for analytical processing. Even then a large portion of MB's inventory is based on low intensity sampling. Although adequate for strategic planning, it has limitations for site specific operational planning. MB manages approximately 1.5 million hectares of land on the coast. The forest inventory is stored in an ARC/INFO GIS that runs on a series of networked Sun workstations. Inventory revisions are run on an annual basis, providing current information for planning. Gradually, non-timber resource information (deer winter ranges, biodiversity corridors, viewsheds, fishery values, etc.) is being added to the system, enhancing the ability of the GIS to provide the information for "total forest" planning.

## Recent initiatives in GIS based strategic planning

There have been two recent initiatives for including spatial resolution in strategic planning. The first, a forest-level approach, has been valuable to MB for introducing concepts and procedures and for providing a data structure useful for our spatial planning needs. It is discussed below under the heading "Forest Level Spatial Data Structure". The second initiative, run at a sub-forest unit level (e.g. a drainage), focused on the impacts of spatial constraints on harvest schedules. It is discussed under the heading, "Drainage Level Spatial Constraints".

### Forest level spatial data structure

New Brunswick research and development efforts (Jordan and Erdle, 1989; Baskent and Jordan, 1991) provided ample evidence that a GIS-based approach to forest modelling, by maintaining geographic distribution of timber inventory, makes it possible to incorporate operational realities such as harvest blocking and leavestrips in strategic planning and to produce prescriptions/schedules in mapped form. Not only is a truer assessment of future wood supply assured but the likelihood of long-term strategic plans being implemented at the operation level is greatly improved.

The spatial model developed in New Brunswick, however, turned out not to be directly applicable to MB's forest circumstance. The model is stand-based, and forms and schedules harvest blocks by collecting neighboring stands into economically harvestable concentrations of volume (see Figure 1). Built around the ARC/INFO GIS the technique relies heavily upon the topological relationships created by this vector-based GIS and works best with highly fragmented forests, i.e., ones with small stands. While this model would be applicable for MB's second-growth areas, it is inappropriate for a forest that also contains extensive areas of uniform timber type, like MB's old-growth areas. In this latter circumstance, stands would have to be partitioned in order to form harvest blocks and leavestrips. Techniques for doing this with vector-based GIS would be tedious at best, and impractical at worst. This necessitated development of a new spatial modelling technique at MB.

As an alternative, the company turned to a raster (or grid) data structure. In simplest terms, a raster data structure would capture the geographic distribution of forest conditions by dividing the entire forest up into a regular grid of cells of specified size (dependent on desired resolution). Beyond the inherent simplicity of the raster data structure, its flexibility in forest modelling is significant. By providing a spatial resolution finer than stratum or stand-level detail, it effectively integrates the different spatial resolution of inventory (stand based) and management prescriptions (e.g. harvest blocks). With equal ease, aggregation and partitioning of forest conditions in strategy testing and evaluation is possible (see Figure 2). By permitting this level of location-specific resolution a raster-based approach promises to improve the quality of information exchange between strategic and operational planners.

Since the ARC/INFO GIS lacks a raster processing capability, MB technologists had to "simulate" a raster data structure with ARC/INFO point coverages. The point coverages are created by first generating a regular grid of points and then executing a point-in-polygon overlay with stand-based forest inventory coverages. Each point in the resultant point coverage is assigned stand type and development information as well as a series of empty attributes for bookkeeping during simulation and for recording intervention schedules that result.

As more layers of data are added to the GIS (non-timber resources, harvesting systems, etc.) the attribute list for the underlying grid cells will expand but the concepts for creating and processing them will remain the same.

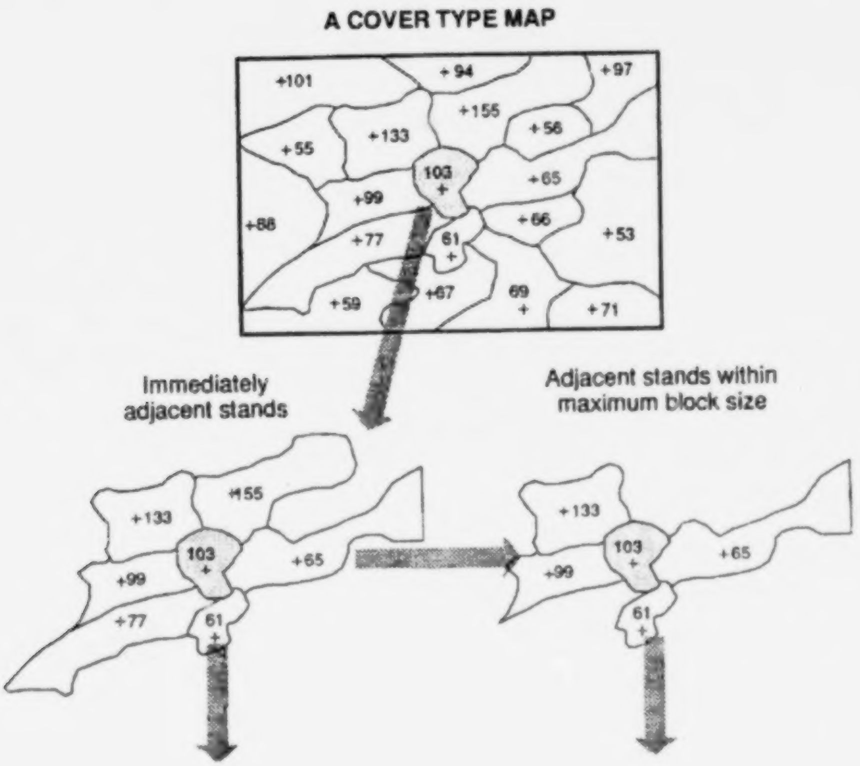
Map results, such as harvest and planting schedules, may be generated following a simulation run by executing an ARC/INFO procedure.

### Drainage level spatial constraints

In a separate initiative FEM has been adapted to examine a number of spatial harvest scheduling constraints. Several drainages under MB management have received a lot of attention and planning effort because of concern about non timber values. Mapped harvest blocks and roads were added for one of these drainages, providing a set of test data for development of planning procedures.

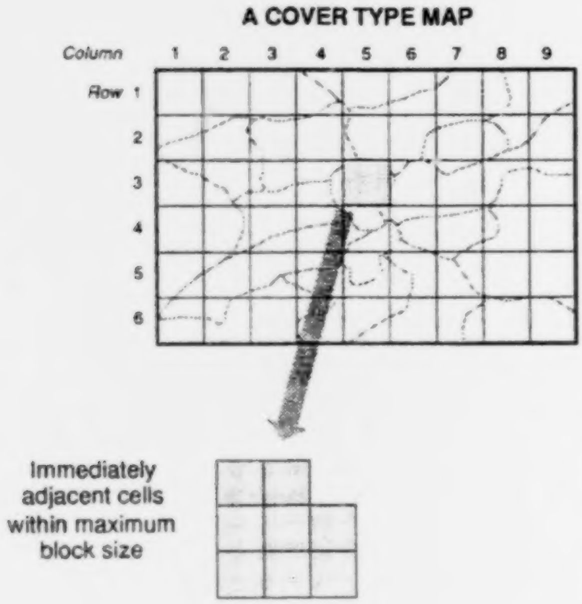
This development has been vector based. Basic polygons are parts of stands divided according to overlapping harvest blocks and non timber concerns such as viewsheds, biodiversity corridors and harvesting the profile. Each basic polygon's description includes identification of harvest block, viewshed, etc.

A number of overlapping timber harvest constraints are recognized. These include adjacency, rate-of-cut and harvesting the profile. A common adjacency constraint is leavestrips (greenup) in which areas adjacent to a


A COVER TYPE MAP
+101
+94
+97
+155
+56
+133
+55
103 +
+65
+99
+88
+66
+77
61 +
+53
+59
+67
69 +
+71
Immediately adjacent stands
Adjacent stands within maximum block size
+133
+155
103 +
+65
+99
+77
61 +
+133
103 +
+65
+99
61 +

| Area (ha) | Operable Volume (m³) | Adjacent Stands | Area (ha) | Operable Volume (m³) |
|---|---|---|---|---|
| 10 | 1 000 | 103 | 10 | 1 000 |
| 25 | 3 000 | 133 | 25 | 3 000 |
| 20 | 2 500 | 99 | 20 | 2 500 |
| 20 | 2 000 | 65 | 20 | 2 000 |
| 30 | 3 500 | 155 | - | - |
| 5 | 1 000 | 61 | 5 | 1 000 |
| 25 | 3 500 | 77 | - | - |
| 135 | 16 500 | <-- Total --> | 80 | 9 500 |

**Figure 1.** Using the topological relationships maintained by a vector-based GIS, a spatial model is able to group immediately adjacent stands into "pseudo-blocks" of maximum size (from Baskent and Jordan, 1991). In this example the maximum block size is 80 hectares. In subsequent scheduling, the same topological information may be used to prevent harvest of adjacent stands for a period of time sufficient to permit greenup.


A COVER TYPE MAP
Column
1
2
3
4
5
6
7
8
9
Row 1
2
3
4
5
6
Immediately adjacent cells within maximum block size

| Area (ha) | Operable Volume (m³) | Adjacent Cells |
|---|---|---|
| 10 | 1 200 | 2,4 |
| 10 | 1 200 | 2,5 |
| 10 | 1 200 | 3,4 |
| **10** | **1 000** | **3,5** |
| 10 | 1 000 | 3,6 |
| 10 | 1 300 | 4,4 |
| 10 | 1 500 | 4,5 |
| 10 | 1 300 | 4,6 |
| 80 | 9 700 | Total |

**Figure 2.** Using a raster-based GIS, a spatial model forms "pseudo-blocks" by grouping adjacent cells, orthogonally as well as diagonally, until a maximum block size (80 hectares in this example) is reached. In subsequent scheduling, harvesting of adjacent cells (and cells up to some minimum distance) may be prevented for some period of time to permit greenup.

harvested area cannot be harvested until the initial area has greened up. The greenup period is often between 8 and 15 years depending on objective and site index. To plan a harvest schedule with greenup constraints it is necessary to know which harvest blocks are adjacent to one another. Adjacency information is developed within the GIS and passed to the Forest Level model.

The rate-of-cut (harvest) may be restricted on a local area basis to protect visual values or for hydrologic concerns. For example, a drainage is divided into a number of viewsheds which vary in visual quality objectives and corresponding rate-of-cut constraints. For each area the constraint is defined as the maximum percentage of area below a defined age.

The intent of a harvest-the-profile constraint is to match in some way the harvest description with the inventory description. For example, there is a desire to harvest according to the distribution of areas by harvesting system within the inventory. If ten percent of the inventory is mapped for long-line systems then it is required that ten percent of the harvest area over a ten-year period is also by long-line systems. To examine the rate of cut and profile constraints it is necessary to identify constraint areas (e.g., viewshed and profile component) for each basic polygon. Again, this is developed in the GIS and passed in the data file to the model.

Initially the main concern was to determine whether there was a feasible harvest schedule given the overlapping constraints. The second question was what impact do different combinations of constraints have on possible harvest levels? The procedures proved useful, showing for the test case, the existence of possible harvestschedules and the harvest volume impacts of the constraints. Further, harvest schedules are readily mapped by passing the harvest output file back to the GIS. Mapped output assists with checking of reasonableness and comparing options. The GIS data base and adjacency information is being used to improve the link between operational reality and strategic planning.

The next stage, underway now, is to add more engineering rationale and to add economics to the harvest scheduling process. In harvest volume terms there may be a number of feasible schedules. Which is preferred? Also, what are the trade offs between planning scenarios? The first question is important to forest engineers. The second is important input to planning policy.

Even in this situation of defined harvest blocks and adjacency relationships the usefulness of a raster data base structure was evident. Many planning problems may be solved by partitioning polygons while retaining locational identity. For example, a small viewshed includes part of a single harvest block. If the block is harvested all at once the viewshed rate of cut constraint is broken. Dividing the harvest block into two smaller harvest blocks may provide an acceptable harvesting alternative while retaining the visual values. This is most easily accomplished with a system of grid cells.

## Where to next?

The spatial constraint accounting system and the raster data structure will be brought together. This will be facilitated in the latest version of ARC/INFO by a noteworthy new raster processing capability (called GRID) based upon Dana Tomlin's map algebra (Tomlin, 1990).

Procedures have been developed on relatively small forest areas. It is necessary to expand the results and the analysis to major blocks (or working circles) within a management unit. For some applications it may be sufficient to use a number of drainage analyses to provide a basis for indirect rules in the larger analysis (similar to the additional inventory requirement in FEM mentioned above). The advantage is reduced complexity in the larger analysis. For other analyses, a more direct expansion of spatial modelling to larger areas is required. Of particular importance is the need for such tools to objectively compare forest management scenarios for policy analysis. What are the trade offs in terms of timber values and non-timber values? Does another option make more sense?

Spatial modelling has the potential to greatly assist the process of formulating twenty-year plans. Data concerns are less limiting than for short-term operational planning where site specific detail is important. The modelling process provides a framework for defining harvest constraints and attaching financial information. This is used to identify feasible harvest schedules and to compare options. Further improvements to this planning process include outputting mapped schedules and encouraging engineers to edit and feed them back to the analysis. These relatively detailed twenty-year plans will be included as the first portion of longer term forest planning analyses to ensure consistency between the two time frames.

Mapped harvest blocks and roads were available in the test area used for the analysis of spatial harvest scheduling constraints. This level of planning detail is not available for all areas. The raster-based data structure provides the flexibility required to develop procedures for simulating harvest blocks by aggregating and/or partitioning stands. Harvest blocks produced in this way may then be mapped to obtain feedback from forest engineers.

## Summary

Use of GIS based strategic forest planning is essential for MB to meet the challenges of managing the forest for a wide range of forest values and for planning the transition from the old-growth forest to the second-growth forest.

In particular, recognition of site specific locations is required to examine the impacts of spatial constraints, provide flexibility for handling different spatial resolutions (e.g., stands and harvest blocks), and to produce maps. These features improve the links between operational planning and strategic planning, communicating operational reality to strategic planning, and a better awareness of strategic plans to operations.

The tools for effective forest-level spatial modelling are coming together. MB has a GIS-based forest inventory for its management land base. Recent work has shown that a raster data structure is most useful for our spatial planning needs. Routines are being developed to examine the impacts of location specific constraints on harvest schedules. The next task is bringing this work together and expanding its use to larger geographic units.

## References

Baskent, E.Z. and G.A. Jordan. (1991). **Spatial Wood Supply Simulation Modelling**. Forestry Chronicle in Press.

Jordan, G.A. and T.A. Erdle. (1989). Forest Management and GIS: What have we learned in New Brunswick? **The Canadian Institute of Surveying and Mapping Journal**, 43(3): 287-295.

Tomlin, C.D. (1990). **Geographic Information System and Cartographic Modelling**. Prentice Hall, Englewood Cliffs, New Jersey 07632. pp. 249

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Elements of management for tracking forest planning in Quebec with the aid of GIS

**Gilles Arseneau**
**Kim Lowell**

M.S. Student and Professor, respectively
Universite Laval
Industrial Chair in Geomatics
Faculty of Forestry and Geomatics
Pavillon Louis-Jacques Casault
Cite Universitaire, Ste-Foy
Quebec, CANADA G1K 7P4
Office: (418) 656-7998
Fax: (418) 656-7411

## Abstract

*(This is an English version of a paper written in French for presentation (in French) at the Canadian Conference on GIS to be held in Ottawa, Ontario in March 1992.)*

*One must optimize the use of GIS to perfect one's knowledge of a forest territory. By implementing a spatial-temporal model to track forestry operations, one would like to create a tool to encourage the use of time in forest management and minimize unexpected surprises when actual operations are realized. Presently, there is very little integration of information in the planning process for forestry operations and exploitation. However, the management of this information is now possible with greater ease. During planning, one must identify how to integrate in the decision process the impacts of forest interventions on the ecosystem. To better exploit the potential of a GIS, it is necessary to store the pertinent management information that constitutes the knowledge base. In summarizing descriptive information about forest management operations, we can increase the precision of our knowledge about the dynamics of the forest.*

## Introduction

With the advancement of a technology such as GIS, we have reached a point where it is necessary to question the practices of an earlier age of forest management. The capacity for management and interpretation of forestry data can now be more easily realized.

Presently, we must develop and integrate a variety of forestry information required for the planning, control, and tracking of forest activities. The problem is more critical now that forest resources of high quality are rarer and the management of the mixed forest, as is the case in Quebec, is more difficult.

Presently, strategic forest planning is conducted with a data base that corresponds to a tactical level — that is, a relatively coarse knowledge of the terrain. It is our intention to determine if one can modify one's planning methods by utilizing more exhaustive data obtained from previous years. We want to conceive a tool to guide planning and to encourage the use of time-based information in order to minimize the effects of surprises.

One must wonder why we do not compare the volumes forecast for a territory with the volume actually obtained for a given territory. These data, considered with a certain amount of human interpretation, would provide an inventory that would allow us to improve our projections of forest resources. Furthermore, one could utilize these data to model better the reaction of the ecosystem to a given operation. However, our present planning methodology does not permit us to establish the necessary procedures for validation and feedback.

Between the planning and realization of forestry operations, there are a number of procedures that one must understand and integrate as knowledge. However, it is necessary to know what one wants and how to get it in order to avoid the phenomenon of GIGO (garbage in, garbage out).

## Factors of management

One of the principal tasks inherent in forest management planning is *monitoring the evolution of the forest.* In the province of Quebec, one utilizes the software SYLVA as a tool to simulate forest yield. SYLVA permits one to define yield objectives and the breadth of silvicultural operations for each beneficiary of a forest management contract. However, existing planning techniques make this process difficult. It is necessary to obtain feedback concerning the annual harvest that can be conducted without reducing the forest capital. Moreover, this is necessary for all levels of forest planning.

It is necessary *to protect commercial forest stands* against the takeover of the forest by undesirable species, maximize the expansion of commercial stands, and accelerate their maturation through a cycle of natural regeneration. The present strategy of forest conservation to achieve these goals must utilize integrated management which considers all forest resources. Furthermore, one would like to conduct forestry operations by minimizing forest interventions and maintaining production. One would also like to eliminate chemical interventions to protect further the environment.

*To minimize forest losses*, we must protect these resources against catastrophes — namely, insect infestations and fires. In knowing all of the factors which favor these, we must prioritize the most susceptible zones during planning. Specifically, considerable efforts must be expended to minimize the effects of spruce budworm. Finally, *to assure adequate wood supplies for the forest industry,* we must alter the characteristics of forest stands as a function of expected future industrial needs.

All of these points contribute to promote the implementation of a spatial-temporal model to study the relation between cause and effect.

The conception of such a model has been conducted within the context of supply for a sawlog and veneer mill. This has evolved with the notion of the scarceness of the resource. Furthermore, the sectors of supply for this are the area of mixed-species forest in Quebec. We have had a tendency to manage these zones for either hardwood or softwood. Currently, we want to regard this ecosystem as a different entity — one which is truly mixed. However, this cannot be done without complicating the work of the forest manager.

## Methods of planning

Before using Geomatics in forest planning and installing a valid spatial-temporal system, it is necessary to place things within a certain context and to understand better the dynamics of forest planning.

In the province of Quebec, we have divided planning into three steps: the General Plan (for 25 years), the Five-year Plan, and the Annual Plan. For each of these, the data base is the one supplied by the Minister of Forests (MFO) (Figure 1). Each of the planning steps correspond to a Strategic, Tactical, or Operational (respectively) planning level.

A problem is that the forest inventory data from the MFO correspond to the Strategic or Tactical level only, and is not suitable for Operational planning. This is the reason that forest management must be conducted on forest inventories obtained by operation interventions in order to validate projections made before operations commence.

## Classification processes

### Stratification

It is necessary now to separate the steps of forest management to obtain an Annual Plan closer to reality than the General Plan. The base of all information systems is the reliability of available data. One must, therefore, validate the information on a forest map using ground control points. After one has obtained a certain confidence in the information contained on a map, one moves into the phase of the classification of forest strata. We have identified four general classes of stratum (Figure 2).

#### *1. Inaccessible strata*

Within the data base of the MFO, there is a criteria of accessibility (which is based on topographic slope) assigned to each forest stand. However, one must also identify stata located in "inaccessible places" such as ridges, bottomlands, and isolated stands. This identification can be conducted by overlaying topography and elevation information.

#### *2. Priority strata*

These are those strata which are susceptible to significant losses of volume if not harvested "immediately." Principally, these include those areas which have been recently subjected to fires or insect infestations,

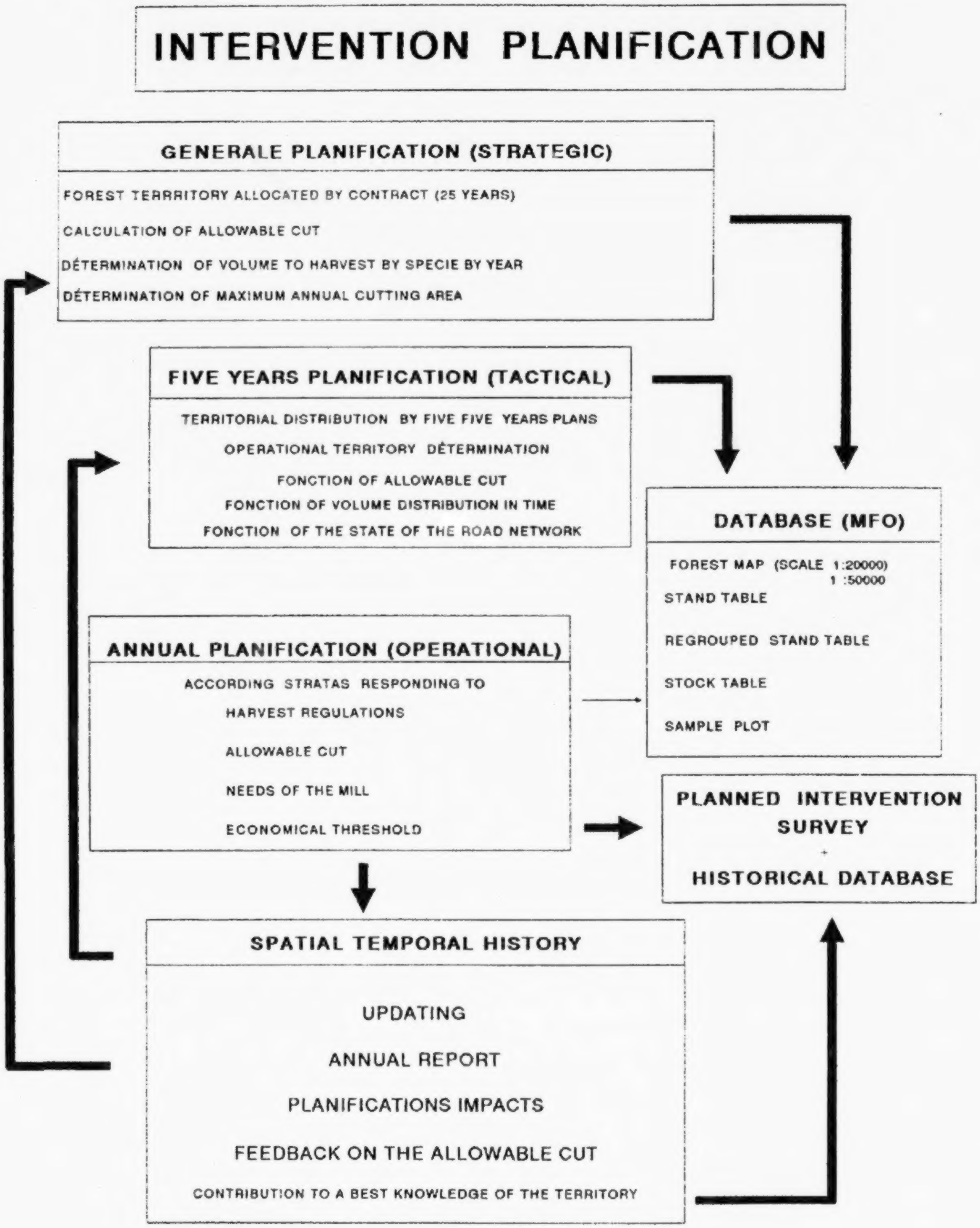

INTERVENTION PLANIFICATION
GENERALE PLANIFICATION (STRATEGIC)
FOREST TERRRITORY ALLOCATED BY CONTRACT (25 YEARS)
CALCULATION OF ALLOWABLE CUT
DÉTERMINATION OF VOLUME TO HARVEST BY SPECIE BY YEAR
DÉTERMINATION OF MAXIMUM ANNUAL CUTTING AREA
FIVE YEARS PLANIFICATION (TACTICAL)
TERRITORIAL DISTRIBUTION BY FIVE FIVE YEARS PLANS
OPERATIONAL TERRITORY DÉTERMINATION
FONCTION OF ALLOWABLE CUT
FONCTION OF VOLUME DISTRIBUTION IN TIME
FONCTION OF THE STATE OF THE ROAD NETWORK
DATABASE (MFO)
FOREST MAP (SCALE 1:20000)
1 :50000
STAND TABLE
REGROUPED STAND TABLE
STOCK TABLE
SAMPLE PLOT
ANNUAL PLANIFICATION (OPERATIONAL)
ACCORDING STRATAS RESPONDING TO
HARVEST REGULATIONS
ALLOWABLE CUT
NEEDS OF THE MILL
ECONOMICAL THRESHOLD
PLANNED INTERVENTION SURVEY
+
HISTORICAL DATABASE
SPATIAL TEMPORAL HISTORY
UPDATING
ANNUAL REPORT
PLANIFICATIONS IMPACTS
FEEDBACK ON THE ALLOWABLE CUT
CONTRIBUTION TO A BEST KNOWLEDGE OF THE TERRITORY

FIGURE 1

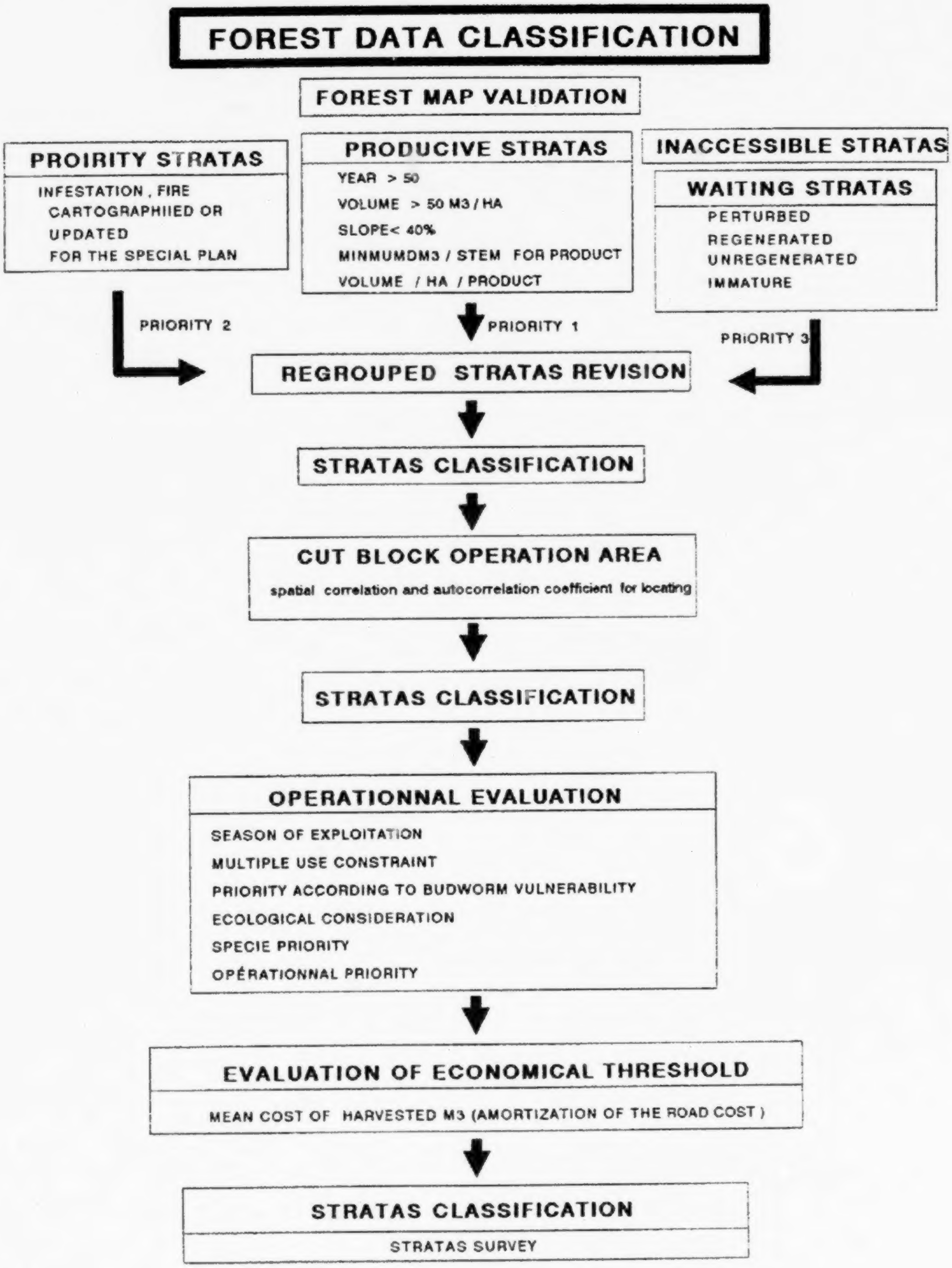

FOREST DATA CLASSIFICATION
FOREST MAP VALIDATION
PROIRITY STRATAS
INFESTATION , FIRE
CARTOGRAPHIIED OR
UPDATED
FOR THE SPECIAL PLAN
PRODUCIVE STRATAS
YEAR > 50
VOLUME > 50 M3 / HA
SLOPE< 40%
MINMUMDM3 / STEM FOR PRODUCT
VOLUME / HA / PRODUCT
INACCESSIBLE STRATAS
WAITING STRATAS
PERTURBED
REGENERATED
UNREGENERATED
IMMATURE
PRIORITY 2
PRIORITY 1
PRIORITY 3
REGROUPED STRATAS REVISION
STRATAS CLASSIFICATION
CUT BLOCK OPERATION AREA
spatial correlation and autocorrelation coefficient for locating
STRATAS CLASSIFICATION
OPERATIONNAL EVALUATION
SEASON OF EXPLOITATION
MULTIPLE USE CONSTRAINT
PRIORITY ACCORDING TO BUDWORM VULNERABILITY
ECOLOGICAL CONSIDERATION
SPECIE PRIORITY
OPÉRATIONNAL PRIORITY
EVALUATION OF ECONOMICAL THRESHOLD
MEAN COST OF HARVESTED M3 (AMORTIZATION OF THE ROAD COST )
STRATAS CLASSIFICATION
STRATAS SURVEY

FIGURE 2

and where recovery of the area is possible. Secondarily, one must identify those areas which are vulnerable to such maladies based on the species, age, and density of thea forest stand. Ultimately, it would be useful to utilize ecophysiographic information to estimate these parameters. Work is ongoing in the MFO to classify the criteria for vulnerability; results should be known soon. One must then try to integrate these, as much as is possible, into forest planning.

***3. Strata in waiting***

***Regenerating strata***

These are those strata which have been subject to (human and/or natural) disturbance and which are regenerating withy commercial species. Reforestation work has been effective to eliminate shortcomings in natural regeneration.

***Young strata***

These are those strata for which the volume and age are not yet sufficient for optimum development. One assigns to them a code indicating the number of years expected before they reach commercial maturity.

***Disturbed strata with volume***

These are those strata which have been subject to human and/or natural disturbance in the past and whose present forest parameters do not conform to any standard intervention. It is possible that, if no corrective action is taken within these strata, it will require considerable time before they reach an acceptable level of productivity. It is thus very important to exploit these territories "in the near future."

***Disturbed strata without volume***

These are those strata which have been subjectto human and/or natural disturbance which have not regenerated with commercial species. It is necessary to be able to restore them to productivity "eventually." These strata constitue an habitual "backlog" of strata that must be restored to production at a relatively high cost.

The Strata in Waiting are those entities for which it is necessary to consider that, if one intervenes within a given sector, these could be integrated into forest planning. The ideal would be to examine the territory with a fine-tooth comb, yet still manage the terrioty in its entirety. However, to optimize investments, it is preferable to choose those Strata in Waiting for which an appropriate intervention would provide wood during a time of scarcity. Also, choices must be based on ecological criteria and experience in order to identify those sectors which will respond well to prescribed interventions. Still, one will determine the ecological cover and compare this with the planning history. To do this, however, one must have followed management planning and maintained these historic data.

***4. Productive strata***

These are those strata that provide the wood supply to a factory for the short- and long-term. These are the pivotal strata in the system.

After identifying these strata, we must revise the re-groupment of strata to respond to management needs. The base of all estimates being a sample, we establish our criteria of discrimination on these results. The MFO has already established one regroupment which we must review. The result of this revision will translate into seven classes:

1 - Well-sampled productive strata
2 - Poorly-sampled productive strata
3 - Well-sampled moderately productive strata
4 - Disturbed strata with volume
5 - Distrubed strata without volume which are poorly regenerated
6 - Young strata
7 - Strata in the process of being regenerated

With the stratification process being standardized, one would be able to facilitate management planning with the integration of an expert system. We believe that stratification is a step in which one can define the mechanisms which encourage certainty, and thus form-alize the representation of knowledge and provide an appropriate decision tool.

"Experts" use a collection of operational concepts. Each of these corresponds to a node within their heuristic. Possibly we have all of the elements in hand to standardize and "systematize" this step based on an expert system. However, to do this requires one to build a data base in which one can have confidence.

## Integration within the "massive forest"

Following this work, one is in a position to start the work in which a GIS is able to help delineate the "Massive Forest" — i.e., relatively large blocks of forest which are able to sustain a harvest. By using a spatial auto-correlation coefficient, we may be able to determine a territorial division (according to a Quad-Tree method) in which certain phenomena — i.e., forest volume — have a tendency to regroup themselves (Figure 3). Instead of examining criteria of contiguity, one would be able to consider minimum distance between forest stands as the criteria of "proximity."

Furthermore, points on the map could correspond to the centroid of polygons (forest stands), and the value at each point could be considered to be representative of the are and volume corresponding to the polygon.

The area of Massive Forests must consider the Annual Allowable Harvest. It will also be necessary to consider this in order to further research in spatial operations in forest planning, and integrate these into the planning process.

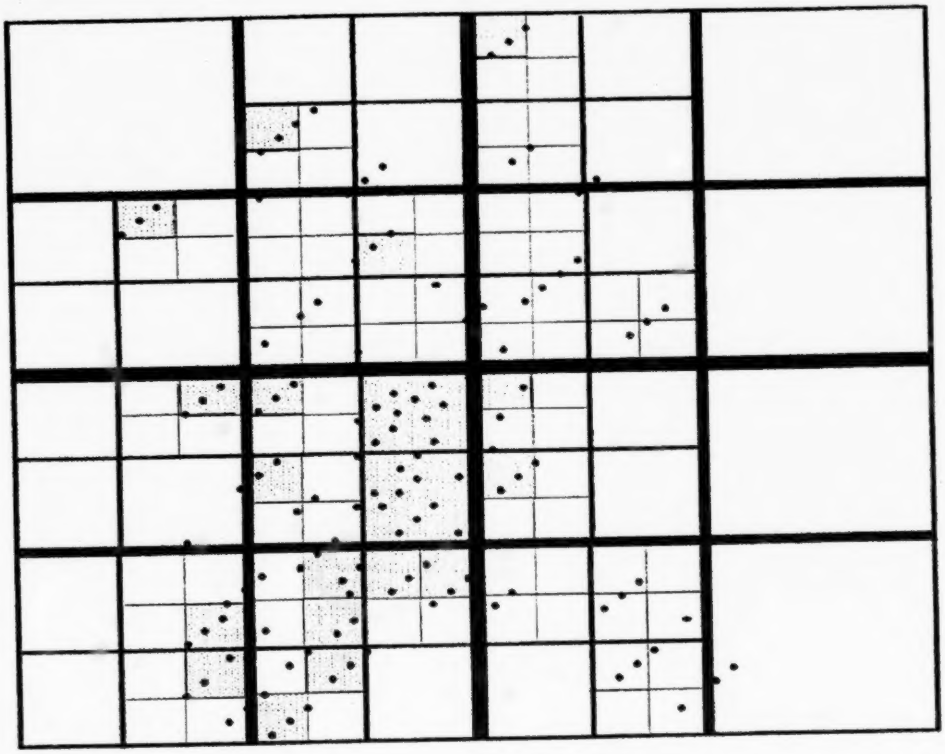


***Figure 3:*** *Schema representing spatial research according to the quad-tree method for locating operational cut block with the spatial correlation and autocorrelation coefficient. It indicates where the phenomena tends to regroupe.*

## Operational evaluation

In this step, we take the results of the identification of the Massive Forest and begin considering biological and legal constraints. To do this, we must plan forest operations in space and time.

For those things constrained spatially, we must submit to the legal normes of forest intervention such as:

- Constraints on exploitability based on riparian zones and site fragility
- Constraints for wildlife habitat
- Constraints based on local population centers
- Constraints to prevent catastrophes

We must put in place mechanisms to analyze cause and effect and to minimize the risks when different alternatives are considered.

To utilize a spatial-temporal system to follow the work of a forest planner, it is possible to make decisions which account for results obtained under "similar conditions." However, to do this one must know all of the prevalent conditions in an area before a decision can be made. In employing such a spatial-temporal system and including relevant historical data, we would be able to decrease the amount of uncertainty in the decision-making process. Foresters operate on ecosystems which react specifically to the actions undertaken during specific conditions. It is left to the forest planner to understand (as much as possible) these reactions.

A spatial-temporal system for monitoring forest operations would have direct effects on the evaluation of operations, and would constitute the heart of decisional processes of the forest manager. It is this system that would help to identify the incompatible methods of planning which are the heuristics of the forest manager. That is, it is in the context of complex forest management that the forest manager must "have a feel." In the province of Quebec, we are in the process of producing ecophysiographic maps which identify the prevailing physiographic conditions within an area as a function of forest stand, drainage class, and slope. We would like to introduce the idea of the "forest station" which integrates the information of cartographic classes to be allow for the management of forest stands based on their "natural tendencies."

Despite the technology which surrounds us, it is necessary to keep in mind that it is the forest manager who ultimately makes the decisions. One should make a list of planning priorities to be able to acknowledge this constantly. Flexibility is essential in any management tool since a seemingly infinite number of decisions must be made. These should be based on judgement and experience in addition to quantitative planning tools.

## Evaluation of economic return

Economic simulation software which exists currently has been created largely to respond to planning needs at the Strategic and Tactical (25- year and 5-year plans) levels. However, in integrating this type of software in the final step of the decision process, we can introduce data which are reliable at the Operational level. Although these data were not designed for this level, it is possible to modify them for this purpose. In the province of Quebec, we have one software system for forest management — Eco-4 — which is able to study the effects of different choices. One can give forestry and economic factors equal weight in Eco-4.

This software examines various scenarios by forest stand and calculates the optimal year for harvesting. It will also calculate the Annual Allowable Harvest of various products according to criteria which have been established to mimimize growth and volume losses.

Furthermore, Eco-4 was designed to provide feedback on a variety of scenarios given the Annual Allowable Harvest. We are thus able to optimize the scenario selected so as to satisfy both the biophysical and economic constraints in an optimal fashion. Finally, Eco-4 is a tool that retains the four dimensions of forest management:

- Forestry practices
- Economic Considerations
- Space
- Time

Finally, following this last step, we are able to classify cartographic strata by intervention priority. We are thus somewhat in a position to verify the prevailing conditions on a given terrioty before harvest.

## Integration of spatial-temporal monitoring

It is first necessary for us to determine the information that one possesses such as data which already exist within an organization. It is also necessary to examine what information is desired as a result of our spatial-temporal monitoring of management operations.

For the payment of subsidies, the MFO requires us to produce reports on forest operations relative to the forest area which has been treated. Furthermore, for the payment of cutting rights, the volume harvested must be reported by species and product.

One of the first things to do is to make a comparison between the work planned and the actual work conducted. We should be able to determine where any differences have occurred and remedy those problems by updating our planning following the analysis of these results.

It will be necessary to be able to organize forest operations to evaluate the origin of wood volume on the territory. These figures will permit us to question and verify certain estimates provided by the forest inventory. We would also be able — through a rigorous description of conditions prevailing before the operations (site quality, drainage, slope, coefficient of distribution) — rectify our extrapolations and improve the precision of the estimate of the Allowable Annual Harvest. If it is necessary to examine a larger territory than that used to estimate the minimum volume required (to supply the contract beneficiaries), it would be useful to identify the reason(s) for this bias. This would be possible through the implementation of a methodology within the data collected for Operational management planning. This would allow one to determine if the bias resulted from, for example, underestimation of deleterious maladies, territory accessibility, the wood volume itself, or the size of trees for certain forest stands.

By spatially referencing the source of our volume by linking data to the smallest possible cell, we can characterize the forest stands.

Furthermore, the spatial-temporal monitoring of sectors of intervention would require us to return to the forest to study the reaction of the sector to our intervention. It is therefore necessary to insert within our monitoring certain "temporal control points" to validate our predictions. To do this, we will have to construct models for temporal monitoring to identify spatially the control points to establish.

## Conclusion

To institute spatial-temporal monitoring, one would collect information to increase our knowledge of future reactions of a forest ecosystem. To optimise the utilization of this resource, we must "systematize" the monitoring procedures and assemble the heuristics of an "expert manager" of the forest. To do this, we must consider the integration and distribution of the information so that the annual planning is not always done more than once. Some forestry companies in Quebec, equipped with a GIS, are already instituting procedures for the operational monitoring of the softwood forest. This forest, which is managed by large companies, has less variation than the hardwood and/or mixed forest. These latter forests often provide those companies making sawn products or veneer, and which hesitate before investing in GIS. But if one convinces them to fund GIS...

One of the basic precepts of a forestry expert system is the reliability of the data base. One has confidence in an expert system for certain procedures — i.e., those things for which the quality of data can be assured.

## Acknowledgements

The authors would like to express their appreciation for the support of this work by the Forest Development Service of Rexfor. Furthermore, we desire to highlight the financial support of the company Abitibi- Price through a scholarship to the senior author. We also acknowledge the support of the Association of Quebec Forest Industries and the Natural Science and Engineering Research Council of Canada.

## Bibiliography

Baskerville, G.L. (1986). Understanding forest management. Forestry Chronicle 62(4):339-347.

Baskerville, G.L. (1988). Tools to guide our decisions — reliability. Proceedings: Symposium O.N.M.R. and Forest Canada, 170-178.

Baskerville, G.L., and T.G.E. MOORE (1988). Forest information systems that really work. Forestry Chronicle 64(2):136-140.

Brand, D.G., and Moore, T. (1990). Current Canadian research in forest management planning. Proceedings: The Interlacement of Enterprise Management and Forest Management planning in Forest Enterprises, 235- 241.

Davis, M.O. (1984). Anatomy of decision support. Datamation, 201-208.

Ganascia, J.-G. (1985). The conception of expert systems. La Recherche 2:1142-1151 (in French).

Hall, T.H. (1981). Forest management decision making: art or science? Forestry Chronicle 57:136-140.

Lussier, L.-J. (no date). Notebook of presentation of a system of forest management: Eco-4 (in French).

Paquet, D., P. Trudel, G. Gauthier (1991). Report of a working group on Sylva II presented to the committe of the Manual of Forest Management (in French).

Werner, M. (1986). To finance the supply of wood. Allocation to the association of sawlog manufacturers of Quebec (in French).

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Probability-based spatial-temporal modelling for natural forests: sampling techniques, model development, accuracy assessment and residual evaluation

**Kim E. Lowell**

Professor Agrege, Industrial Chair in Geomatics
University Laval
Local 1327, Pavilion Casault
City University, Ste-Foy G1K 7P4
Quebec, Canada
Tel: (418) 656-2530
Fax: (418) 656-7411

### Abstract

*Methodology to produce spatial-temporal models for natural systems using a GIS is presented and discussed. The goal of the models covered is to be able to estimate the future amount and location of land cover types given existing land cover and variables which are related to the natural evolution of the system. Specifically, three subjects are covered in detail — data sampling from a GIS, statistical techniques for use with categorical map data, and the spatial evaluation of the models. The latter includes accuracy estimates as well as an assessment of spatial independence of model residuals.*

## Introduction

In many disciplines, existing spatial quantitative models are being adapted to a geographic information system (GIS) to add a spatial element to prediction. One example is the GIS-based use of wildlife habitat suitability models. Generally, these have been developed to answer questions such as "How much optimal (moose, e.g.) habitat?" exists on an area. The adaptation of such a model to a GIS is indicative of a desire to extend such models to be able to answer questions of "Where is the optimal habitat?" A comparable analogy exists in forestry in the form of growth and yield models. These generally project "how much" wood may exist at some time in the future, but do not indicate "where."

While models such these are useful, a key to all such "adaptive work" is the availability of pre-existing models. However, these generally have not been produced using cognizance of spatial relations and object locations in the development process. Thus their adaptation to spatial systems is fraught with hidden perils. Principal among these is that, generally, these models have been developed from a relatively small sample of data and little attention has been paid to micro-site differences and the spatial relationship of a site with its neighboring sites.

A GIS is a data base which contains this type of locational information for a given area. A GIS has the potential to be the principal source of analytical data for the development of predictive models. With existing aspatial models, the "answer" to a query is generally a single number ("the amount of....") and possibly its associated confidence limits. For the spatial models to be discussed in this paper, "the answer" that is desired is a map that can be summarized if necessary to answer "How much?" Of specific interest in this work is the development of predictive models which utilize present landcover and related variables — e.g., elevation and soil — to project present landcover to some time in the future. Furthermore, it is desired to develop such models using the maps contained in a GIS as the sole source of data input.

Models such as these have a myriad of uses. One is to explore the spatial and aspatial dynamics of a natural system to determine which factors relate to changes in the system. Another purpose might be to use a model which has been developed from existing historical data to make inferences about future conditions on a natural system. And, finally, one might develop a model for an area for which one has "a lot" of data and be able to apply it to an area for which information on the dependent variable is limited (provided, of course, that conditions on the two are similar).

The purpose of this paper is to present a generalized methodology for developing from a GIS spatial models which project the future location and amount of land cover types. In doing so, specific examples will be given from previous work by this author and others. Three issues will be covered in separate sections: 1)how to sample a GIS to obtain data for model development, 2)statistical techniques for use with categorical map data, and 3)how to evaluate such models.

## Sampling a GIS

What is meant by "Sampling a GIS" is that one must somehow produce an (aspatial) ASCII file for subsequent analysis by a statistical software package. Thus the goal of this section is to describe possible methods for imposing some sampling system on an existing GIS in order to produce an ASCII file for which each "record" has all the information available in the GIS for a selected location/entity. In existing GIS, a location is generally thought of as a dimensionless point or raster cell in existing GIS whereas an entity is a polygon.

If one is using a vector GIS and thus implicitly assuming that an individual polygon is the basic unit of interest, sampling is relatively straight-forward. Because the polygon is treated as being completely uniform within (relative to the variable being mapped), one need only access the "attribute file" associated with the forest map and output one record for each polygon. Such a situation might exist in the case of plantation forestry in which a stand (polygon) is planted, treated, and harvested as a unit — the stand is never sub-divided and stand boundaries do not change. If one wishes to relate, for example, plantation growth to soil type and topographic aspect in such a situation, one imposes the boundaries of each map on a base map of common scale. Topology is then rebuilt to redefine polygons which are, again, being considered to be uniform within. One record per polygon can then be exported from this (overlain) polygon map to produce the necessary ASCII file.

In a raster GIS, the situation is much more complex. For the work presented herein, a raster must be viewed as a regularly-spaced sample of the underlying landscape. On that landscape, it is assumed that polygons exist which are "meaningful" for a parameter of interest. However, the composition and boundaries of these polygons are assumed to be unknown a priori. For example, one may have maps of two factors (A and B) which may be related to factor C. If A and B are both important, one overlays the boundaries of both features to identify the underlying polygons. If only A is important, however, only the boundaries of classes on A would define the underlying polygons. It might also be the case that maps of A and B are both important, but not all classes on A. Thus one would have to aggregate some classes in A and use the boundaries of these with all of those on B to define the underlying polygons. (In a raster GIS, the "polygons" discussed are groups of contiguous cells which are recoded to have the same value.)

The preceding discussion makes it evident that, even for only two maps, there is a large set of possible polygons which can be defined. Thus one must sample the GIS in such a manner to ensure that all potentially important combinations of factors are "adequately" represented in the data base. That is, one must sample in order to be sufficiently representative of these definite but a priori unidentifiable polygons. In this sense, one may view the eventual model development as the identification of factors for the definition of polygon boundaries.

Because of this approach, the sample density must be guided by the amount and variability of each map in the GIS. For example, in one previous study (Lowell 1989), it was desired to project future land cover based largely on whether or not a fire had previously occurred at a particular locale. Because only 6% of the entire study area had been burned, relatively few sample locations were available for burned areas. Thus, 100% of the burned cell locations were selected, whereas only 10% of the unburned areas were (randomly) sampled. (This produced 7800 and 4953 sample cells for unburned and burned areas, respectively.) Using standard GIS overlay functions, the soil, initial cover type, topographic aspect, distance to a drainage, elevation and landcover for 7 years (over a 48-year period) was subsequently determined for each sample cell and exported to an ASCII file which was then analyzed using SAS (SAS 1979).

In another study (Lowell 1991), it was desired to model landcover from 1939 to 1982 on a 932 hectare wildlife refuge area in Missouri (United States).

There were four landcovers of interest — Forest, Cedar, Open, and Disturbed. On the area, there was "a lot" of Forest, Disturbed, and Open and very little (7 ha) Cedar. The variability in ecological succession for Disturbed (a man-made type) was known to be zero (0); Forest was also expected to exhibit low successional variability over time. (That is, Forest would remain Forest in the absence of disturbance.) Thus these cover types were sampled systematically using a relatively sparse grid — 1 point per 1.5625 ha for Disturbed and 1 point per 2.25 ha for Forest yielding a total of 90 and 223 sample locations, respectively, for each. Open was expected to be extremely variable relative to ecological succession and was sampled at a rate of 1 point per 0.36 ha yielding a total of 782 points. (Open areas would yield to Cedar and then Forest with conversion rates varying greatly over the area.) Cedar was expected to be moderately variable, but the low amount of the type necessitated a high density sampling grid to obtain "enough" sample points. Thus Cedar was sampled at a rate of 1 point per 0.16 ha for a total of 46 cells. At each sample location, soil, aspect, distance from a Forest seed source, and landcover for each of the 7 years was sampled and the information exported to an ASCII file.

These examples are not presented to suggest that these are "ideals" in how to conduct this work. Rather they are indicative of the type of considerations one must utilize in sampling from a GIS for statistical analysis. As in aspatial work, the sample size and the manner in which samples are chosen must be guided by the data itself, and the desired use to which it will be put. In the studies cited, random and systematic sampling were used as was a complete census of one type in one case. For reasons germane to a each study, the method chosen for obtaining data was deemed optimal.

One may question why one bothers to sample from a GIS at all. After all, a GIS might be considered a census of all locations on a study area. As such, why not merely examine every cell or polygon rather than a sample? One must remember that while a GIS might be regarded as an areal census, there are numerous "real-world-to-map" simplifications inherent in the physical maps which are stored in a GIS. Thus, whereas an aspatial population census will produce a standard error of zero (0), consideration of a map-based GIS as a comparable spatial census is grossly misleading. Using all cells/polygons in a GIS to develop predictive models will not allow one to diminish the amounts and effects of cartographic simplifications on the subsequent analysis. Thus there remains a sampling error that cannot be eliminated by merely using all cells/polygons in a GIS. Thus, once a sample size exceeds a certain point, there is little advantage in selecting additional points.

## Statistical techniques for categorical map data

All data stored on conventional thematic maps are categorical. Even contour lines on a topographic map can be thought of as polygon delimiters rather than lines of known elevation. Thus there is limited use for analytical statistical techniques which utilize a continuous dependent variable when working with data from a GIS. One exception is when one has a GIS composed of polygons whose boundaries do not change over time and for which continuous attributes are associated with each polygon. In studies of human geography, therefore, where polygon boundaries are surveyed "township" lines and one wants to use a GIS to model the change in urbanization, one need model only the change in attributes — mean household income, etc. — of each township. In such cases, well-known techniques such as multiple linear regression may readily be used.

It can be argued that the same situation exists in plantation forestry. Boundaries of plantations do not change, and each plantation polygon is considered completely uniform within. In such a case, "spatial-temporal modelling" may require little more than linking an acceptable growth model — which is based on continuous stand attributes such as basal area and volume — to a GIS. The growth model would change the attributes of each stand appropriately and the GIS would be used with "(volume) per hectare" values to estimate the total (volume) in each stand, or over the entire forest.

One may wonder why unchanging boundaries are important in order for only attribute modelling to be necessary. If boundaries are changing, then so (potentially) are spatial relationships. This means that certain polygons may disappear completely, change (categorical forest) type, and sub-divide and/or reaggregate. In this case, modelling the change in attributes only will probably be so simplistic as to miss the "most important" dynamics of the system.

In natural forests, one cannot consider boundaries to be unchanging. Forest stands often change size and shape and regularly reaggregate with other stands. Harvesting operations may create one uniform stand polygon — a clearcut — from pieces of three diverse stands. Of greater interest here are changes in polygons caused by ecological succession. For example, in the northern hardwoods forest type, pioneer poplar (Populus spp.) is replaced by mid-successional birch (Betula spp.), which in turn yields to sugar maple (Acer saccharum Marsh.). As this occurs, stand "polygons" change form and size in a dynamic fashion. Thus for spatial modelling of natural forests one must be able to model the evolution of polygons over time as well as changes in their attributes. Because this involves being able to project — among other things — when a given portion of a stand polygon changes its (categorical) type, one needs statistical techniques which can employ discrete dependent variables.

One possible technique is logistic regression which uses a binary dependent variable with continuous independent variables. Resulting models estimate the probability (0.00 to 1.00) of the occurrence of the binary event being studied. Lowell (1989) used it with a GIS to produce a spatial model to estimate the probability of forest existing at some future time over a 4985 ha wilderness area in Missouri. A classification technique based on a random number generator was then used to assign cells as "forest" or "not forest." Such an approach has not been confined to forest modelling, however. Bernknopf et al. (1988) used this technique to model the probability of the occurrence of landslides in Cincinatti, Ohio. Nor is work with binary dependent variables confined to logistic regression. Bonham-Carter et al. (1988) used an integrated system of equations derived from multiple linear regression to estimate the probability of gold being present at any given location over a fairly large area.

While such approaches are useful if one has a binary dependent variable, they suffer from a serious problem, namely that most maps have more than two categories. Thus statistical techniques which "accommodate" more than two categories of a dependent variable must be explored. A general heading (and a specific technique) is Canonical Correlation. This is related to a technique called Discriminant Function Analysis (DFA) which is of interest here.

DFA produces models which assign "unknown" objects to a category based on the characteristics of the object. Thus the "answer" from a DFA model is the category to which an object belongs. This technique has been used in, among other subjects, remote sensing classification (Ait-Belaid et al., in press). Of relevance here is that Lowell (1991) used DFA to predict land cover at some point after 1939 over the Missouri wildlife refuge area mentioned earlier. Using the data sampled from the GIS (as described in the previous section), he developed four DFA models to identify the most likely class — Forest, Cedar, Open, Distrubed — to which to assign a raster cell at some point in the future. The difference in the four models was the inclusion/exclusion of various independent variables. Each of the four contained some combination of initial land cover, topographic aspect, and distance from a Forest seed source. They were evaluated as will be discussed in the following section.

The examples of DFA cited in the previous paragraph have used it exclusively as a classification procedure. As such, the assignment of an unknown cell/pixel to a given class was equivalent to assigning an object to its "most likely" class. While such an approach makes intuitive sense, it suffers considerably if a given object is almost equally "near" (in statistical space) the centroids of more than one group, or if it is extremely far from the (statistical) centroids of all groups. Thus instead of deterministic classification, it might be useful instead to utilize a probabilistic approach as is employed with a technique like logistic regression. That is, one would like to examine the probability of group membership for each group rather than merely assigning the unknown cell to its "most likely" category.

A single DFA model is composed of a series of classification function for which the number of functions is equal to the number of categories of the independent variable. To utilize a DFA model for a given cell, one calculates a "discriminant score" for each category's function. If utilized deterministically, one then assigns the cell to the class whose equation yields the largest discriminant score. While these scores have no intrinsic meaning, Klecka (1980) explains how they can be converted to Mahalonobis distances which are measures in "chi-squared units" of the statistical "distance" that a given object is from each category's centroid. Because these distances conform to a chi-squared distribution they can, in turn, be converted to the probability of membership in each category. Thus DFA models can be used to determine the probability of class membership for each cell, rather than merely assigning an unknown cell to a particular class. Lowell (in review (b)) has used this approach for the Missouri wildlife refuge area in an effort to increase the spatial validity of DFA models developed. This will be discussed further in the following section.

In closing this section, a word about probabilistic approaches in general is in order. While the use of probability surfaces or "maps" may show promise for this work, it is acknowledged that it is difficult for humans to comprehend fully such maps. Furthermore, such an approach suggests that one needs a separate probability surface for each category of interest. If many classes exist, this soon may become unwieldy if one is dealing with physical maps. Storage and manipulation of these is not a problem in a computer-based system, however, and a probability approach may be a "more honest" way to conduct spatial analysis using a GIS. In typical GIS analysis, one wishes to identify, for example, the "optimal location" of a factory based on a variety of factors (soil, drainage, geology, stratigrafy, distance from a road, etc.). To do this, overlay analysis is generally used to impose the boundaries of each feature on a base map of common scale, and polygons having all of the required attributes are labelled "acceptable" and all others "unacceptable" based on some model. (Such a model is usually descriptive: "the combination of Soil A, Drainage Class 5,....") In reality, the boundaries of each input map are not exact, and there are varying grades of "acceptability" for each factor. Thus instead of asking of the GIS "Show me all the areas which definitely have the following characteristics" it is more reasonable to ask "Produce a map showing the likelihood of having the following characteristics." This concept has been discussed elsewhere relative to fuzzy set theory from mathematics (Robinson 1988).

In conclusion, probabilistic approaches conceptually lend themselves to the work described and also to "the way the real world operates." In order to make probability approaches acceptable to non-technical planners, however, it will probably always be necessary to be able to produce categorical thematic maps. Furthermore, these concepts have yet to be embodied into GIS software. Thus further work in probabilistic approaches will require the development of a series of techniques for utilizing and classifying the probability surfaces, and possibly new data structures for the storage and retrieval of these surfaces.

## Evaluation of spatial models

One can make two evaluations of spatial models which are fairly straight-forward. Here these are referred to as "Areal" and "Locational" Accuracy. Areal Accuracy is the absolute value of the difference between the total predicted and actual amounts of each type. Thus small values for Areal Accuracy are desirable. For example, if there are known to be 100 ha of Type A on an area at a given time and a spatial-temporal model projects that there will be 75 ha, the Areal Accuracy is ($|75 - 100| =$) 25 ha. Areal Accuracy can be determined from a GIS by simply examining the total amount of each type on the Actual and Predicted maps. Providing such information is a standard function of GIS.

Locational Accuracy measures how much of each type predicted is predicted to be located in the correct place. For example, suppose that there is actually 100 ha of Type A and a model also estimates that there will be 100 ha. The Areal Accuracy is ($|100 - 100| =$) zero (0). But if only 25 ha of the predicted Type A is placed at locations having Type A on the actual map, Locational Accuracy would be ($25 / 100 =$ ) 25%. Locational Accuracy can be determined by overlaying the Actual map with the Predicted map and determining those places where the same type is coincident. This type of overlay is also a basic function of GIS.

As an actual example of these, in the first of a series of studies Lowell (1991) specified and calibrated four (deterministic) DFA models to estimate future landcover. In that study, only Areal Accuracy was examined and was found to be comparable for all four models. That is, the total amount of each type was predicted equally well regardless of the individual model chosen. However, a subsequent study (Lowell, in review (b)) on the same area used a probabilistic approach to the same problem. The probability of a given cell being a type at a given time was calculated from the models and cells were then classified in a manner proportionate to their probabilities (using a random number generator). This latter approach caused Areal Accuracy to improve considerably over the deterministic approach, while Locational Accuracy remained consistent for all models.

The third and final evaluation of spatial models is considerably more complex. An assumption of almost all statistical techniques (including those discussed herein) are that residuals are independently distributed. Spatially, however, it is rare that any observations are independent. Thus to ensure the spatial integrity of model estimates, the spatial dependence of residuals must be measured. The spatial dependence of observations is termed "spatial autocorrelation" (SA) and a number of techniques for measuring SA have been described (Griffith 1987). The basic requirements of all of these techniques is that one be able to determine which spatial units (polygons, usually) are adjacent, and what is the (residual) value for each.

Thus one must be able to identify basic map units and compare the values for each on the Actual and Predicted maps to produce a Residual map. There are effectively three ways to do this (Table 1a). Regardless of method, one must first identify the basic map unit. One way is to use the predictive model a posteriori to define polygons. That is, an independent variable in a model indicates that conditions for that variable differ in some fashion from other conditions. Thus these variables could be used to develop a "unique conditions" map which identifies all categories (and thus polygons) for which the model can produce different values. A second way to define basic map units is to utilize the sample points themselves. Adjacency of points can be determined by generating Theissen polygons around each point and recording which polygons touch. Subsequent "residual attributes" are determined only by the value of the sample point at the heart of the polygon. The mention of Theissen polygons suggests the third way to identify basic map units. Instead of using polygons defined by the model, one would use the entirety of each Theissen polygon generated around a given point. The difference between this and the preceding method is that in this case, "residual attributes" would be determined by examining the entire surface of each Theissen polygon rather than only the point value.

Once the basic map units are defined ("somehow") it remains to determine the residual value. This involves a comparison of the land cover on the Actual and Predicted maps for each map unit. Points are the easiest case as one need only examine the individual location to determine if it is the same on both the Actual and Predicted maps (Yes/No). In such a case, residuals are binary — either a point is predicted correctly, or it is not (Table 1b). Polygons are considerably more difficult, however.

Imagine that polygons have been identified as the basic map unit. It is now desired to determine the residual value for one particular polygon. On the Predicted map, this polygon will have been assigned a single value since the polygon will be uniform relative to the variables in the model. (Even Theissen polygons will be uniform since the predicted value for each sample point will be assigned to its entire Theissen polygon). To determine the residual value for this polygon, one must examine the same area on the Actual map. One could assign a binary "correct/incorrect" label to the polygon by determining if the most prevalent class on the Actual polygon agrees with the Predicted value. This has been termed a Plurality rule (Lowell, in review (a)) because the most prevalent class will not necessarily cover a majority of the polygon on the Actual surface (Table 1b). One could use a Majority rule instead. That is, the polygon is labelled "correct" only if more than 50% of the polygon on the Actual map is covered by the same type as on the Predicted map. A final way to determine the residual value is to use a continuous measure rather than a binary one. In this case — termed the Percent rule — one examines the percent of the Actual polygon which is the same as that predicted and assigns this percentage to the polygon.

Having followed one of these procedures for all map units, one is prepared to measure SA. To do so, one must determine which of the unique polygons — Theissen or otherwise — are connected to each other. Following this, if one is using a binary residual produced from the Plurality or Majority rule, one uses the Join-Count Statistic (see Griffith 1987 and Odland 1988). The Join-Count statistic evaluates whether or not there are a disproportionate number of "Correct-Correct" (CC), "Incorrect-Incorrect" (II), and/or "Correct-Incorrect" (CI) links among polygons which are adjacent. The resulting statistic can be tested using a Student's t distribution to assess statistical significance from zero (0). If there are "too many" CC and/or II links, positive SA, or non-random grouping is evident. If there are "too many" CI links, negative SA or non-random diffusion is indicated.

**Table 1a.** Possible ways to identify map units.

| Basic Map Units | Adjacency Determined by: | Predicted values based on: | Residuals based on: |
|---|---|---|---|
| Model-defined polygons | Polygons | Polygons | Areas |
| Points | Theissen Polygons | Sample point | Points |
| Theissen polygons | Theissen Polygons | Sample point | Areas |

**Table 1b.** Possible decision rules to employ for determining map residuals.

| Spatial Unit | Name of Rule | Type of Residual | Description |
|---|---|---|---|
| Points | Yes/No | Binary | Cover type for sample point agrees (or not) on Actual and Predicted maps |
| Polygon | Plurality | Binary | Cover type on Predicted map agrees (or not) with most common type on Actual map |
| Polygon | Majority | Binary | Cover type on Predicted map agrees (or not) with most common type on Actual map and covers at least 50% of polygon on Actual map |
| Polygon | Percent | Continuous | Residual value is the percent of polygon on Actual map which agrees with cover type on Predicted map |

If one uses the Percent rule, one has continuous residuals measured at the interval/ratio measurement scale. Thus one must measure SA using a technique other than the Join-Count statistic. For this purpose, Geary's Contiguity Ratio and Moran's I are available (see Odland 1988, Griffith 1987). Though they embody slightly different concepts, both assess whether or not there are "too many" polygons of similar or different values which are adjacent. Both can be tested for significance from zero (0) using a Student's t distribution and, for both, a positive value indicates non-random grouping whereas a negative value indicates systematic dispersion.

Regardless of how SA is measured, results must be used to modify models appropriately, or select a single model from a number of alternatives. The presence of positive or negative SA is indicative that the model is not "correctly specified." Positive SA — which tends to be more common — generally indicates that there is a spatially spurious — though possibly statistically significant — variable in a model which might be removed without harming Areal or Locational Accuracy. The presence of negative SA probably indicates that an "important variable" has not been included in the model. Lowell (1991) found that the residuals of a DFA model which utilized three ecophysiographic factors to predict future land cover were replete with positive SA — i.e., grouped residuals. Removal of two of the factors eliminated the positive SA while Areal Accuracy was maintained. (Locational Accuracy was not evaluated in that particular study).

Because many options exist for the definition of polygons and residuals exist, some guidance as to which to use is in order here. Germane to the work presented here, Lowell (in review (a)) has evaluated each possible combination of basic map unit definition and residual determination rule for the wildlife refuge area in Missouri. By using only the sample points and the Yes/No residual rule for the associated Theissen polygon, positive SA was rife on the Residual maps for a "Full Model" and a "Reduced Model". (The Reduced Model eliminated two ecophysoigraphic factors included in the Full Model.) This was not surprising since it was expected that, when the original sample grid was established, neighboring points would probably be part of the same polygon (which were considered undefinable *a priori*). The same held true (i.e., positive SA), and for the same reason, when residuals were assessed using the entire Theissen polygon around each point for residual calculation. This was true regardless of the "residual decision rule" employed — Plurality, Majority, or Percent. Using polygons which were defined by the variables contained in the two models, both the Plurality and Majority rules — which produce binary "correct/incorrect" — residuals gave the same results for both models. Specifically, the residuals of the Full Model were overrun with positive SA whereas the Reduced Model eliminated this phenomenon. In a continuation of all the work mentioned (Lowell, in review (b)) the Percent rule — which produces continuous residuals — gave slightly different results for four models evaluated, but only when Moran's I was used to measure SA. Geary's Contiguity Ratio continued to show excessive positive SA in the residuals of Models 1 (the Full Model), 2, and 3, but was eliminated from Model 4 (the Reduced Model). Moran's I, however, suggested that the residuals of both Models 2 and 3 (as well as Model 4) were not spatially autocorrelated in any fashion. (Areal and Locational Accuracies were nearly identical for all four models.)

It is evident that there is no absolutely correct way to apply these techniques in a similar study elsewhere. As with all statistical techniques, results must be interpreted in a "sensible" way and adjustments made accordingly. In the studies cited, SA was used as a way to choose among alternative models whose Areal and Locational Accuracies were comparable. An alternative which was not discussed here is that it is also possible to fit a spatially autoregressive model containing a variable which is employed to quantify explicitly (or "soak up") the SA which may be present (Cliff and Ord 1973). However, this is only possible if one can define the polygons of interest a priori. In spatial-temporal modelling of natural forests, this can be difficult.

## Conclusions

In spatial-temporal modelling for natural forest stands using a GIS, there are three important aspects: data sampling, statistical techniques, and model evaluation. Data sampling must be conducted based on the goal of the desired models and the form of the data. Statistical tecnniques must be able to utilize categorical data. Logistic regression for two classes and discriminant function analysis for more than two classes show promise for both proabilistic and deterministic approaches. Spatial models must be evaluated relative to their Areal and Locational Accuracies, and the spatial dependence of model residuals. This latter, in particular, must be used to guide model selection and/or revision.

## Acknowledgements

For support of this work, the author gratefully acknowledges the support of the Association of Quebec Forest Industries and the National Science and Engineering Research Council of Canada.

## Bibliography

Ait Belaid, M., Beaulieu, J.M., Edwards, G., Jaton, A., and Thompson, K.P.B, 1991. Post-segmentation classification of images containing small agricultural fields. Geo-Carto (in press).

Berknopf, R.L., Campbell ,R.H., Brookshire, D.S., and Shapiro, C.D., 1988. A probabilistic approach to landslide hazard mapping in Cincinnati, Ohio with applications for economic evaluation. Bulletin of the Association of Engineering Geologists, 25, 39-56.

Bonham-Carter, G.F., Agterberg, F.P., and Wright, D.F., 1988. Integration of geological datasets for gold exploration in Nova Scotia. Photogrammetric Engineering and Remote Sensing, 54, 1585-1592.

Cliff, A.D., and Ord, J.K., 1973. Spatial autocorrelation (London: Pion), 178 pp.

Griffith, D., 1987. Spatial Autocorrelation: A Primer (Washington: Monograph of the American Association of Geographers), 75 pp.

Klecka, W.R., 1980. Discriminant Analysis (Beverly Hills: Sage), 71 pp.

Lowell, K.E., 1989. Spatial and temporal GIS modelling using probabilistic functions. Proceedings: *Challenge for the 1990s: GIS National Conference.* Ottawa, Canada, March, p. 1274-1287.

Lowel, K.E., 1991. Utilizing discriminant function analysis with a geographical information system to model ecological succession spatially. International Journal of GIS, 5, 175-191.

Lowel, K.E., in review (a). Measuring spatial autocorrelation among model residuals: continuous and discrete, points and polygons. International Journal of GIS.

Lowel, K.E., in review (b). Probabilistic spatial-temporal GIS modelling involving more than two map classes. International journal of GIS.

Odland, J. 1988. Spatial Autocorrelation (Beverly Hills: Sage), 87 pp.

Robinson, V.B., 1988. Some implications of fuzzy set theory applied to geographic databases. Computers, Environment, and Urban Systems, 12, 89-98.

Sas Institute, Inc., 1979. SAS User's Guide: Statistics, Version 5 Edition (Cary, North Carolina: SAS Institute, Inc.)

# C1 GIS for wildlife modelling

---

**The Map Machine: using the structured query language to produce precisely tailored GIS maps of telemetry observations**

Norm Cimon et al., USDA Forest Service, La Grande, OR, USA

**Integrating spatial models of recreation, wildlife, and timber resources**

Lloyd Queen, University of Minnesota, St. Paul , MN, USA

**A biophysical process approach to wildlife modelling in forestry**

T.W. Norton et al., The Australian National University, Canberra, ACT, Australia

**BC Indian fisheries information system communication techniques for consultation and negotiation interm report**

Joe Kuhn et al., Department of Indian Affairs and Northern Development, Hull, PQ, Canada

---

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# The map machine: using the structured query language to produce precisely tailored maps of telemetry observations

**Norm Cimon**

Supervisory Computer Systems Analyst
U.S. Forest Service
Forestry and Range Sciences Lab
1401 Gekeler Lane
La Grande, Oregon 97850

**Priscilla Coe**

Oregon Department of Fish and Wildlife
Forestry and Range Sciences Lab
La Grande, Oregon 97850

## Abstract

*The Starkey Project is designed to answer questions about how elk, deer, and cattle partition resources in the Blue Mountains of Oregon and Washington. It uses an automated telemetry system which is Loran based, and capable of reporting one animal position every 15 seconds. To manipulate and display this data we use a relational database management system (DBMS), and GIS. The telemetry data are loaded into the DBMS. Electronic forms are used to specify daily and seasonal windows, a list of animals, and the size of the map pixels on which to count visits. The views of the data resulting from the intersection of these criteria are then available for query. A menu selection allows for the creation of a GIS import-ready file of pixel centers in UTM coordinates, and the number of times that each of the pixels was visited, given the underlying criteria. The map name and the specifications used to create it are stored in the DBMS to meet metadata needs. The resulting maps will form the basis for statistical and visual studies of the animal's movements. Animations have been created to show changes over time, according to whatever animal groupings and time cycles we specify.*

## Introduction

*Project Overview* The problem of manipulating and displaying large geographically based data sets in a reasonable amount of time is at the heart of the wildlife studies which have been planned for the Starkey project.

The project is centered on the 25,065 acre Starkey Experimental Forest and Range (Johnson, 1991). The external area has been encircled with animal-proof fencing. Internally, the northeast corner is separated from the rest of the area with this special fencing. This area will be more intensively managed than the rest of the forest and range. There is also a complex of pathways, pastures and pens which make up the generalized "handling" facility. This facility provides the project with winter feeding capabilities, and with the ability to assess the health of the study animals.

The telemetry system is Loran based, with automatic paging of the collared animals. The paging sequence is arbitrary, and not all of the collared animals have to be paged during a given sequence. A given paging sequence is active until a new sequence replaces it (Dana, 1991).

Real time control and monitoring, as well as off-line storage and review capabilities, are provided for this custom built Loran based telemetry system by two 80286 based micro computers which use the MS-DOS operating system. It is an open question whether this system could have been proposed without the availability of such cheap computing capacity at this relatively remote location. It would certainly have been more complex and more expensive.

As important as developments in hardware have been to the success of the project, however, the relatively recent developments in relational database software are the determining factor in the ability of our small group to manage the large volumes of data being reported out by the system in a useful and timely manner.

## Database management

The theoretical origins of the relational model are due to E. F. Codd (Codd, 1970). These origins are firmly grounded in set theory[1]. This work was followed, over the next 11 years by the development of a formal language, a test database, and, eventually, commercial versions employing the Structured Query Language, or SQL (Date, 1989a).

The SQL language is at the heart of most of the new generation of database management systems on all classes of computer systems. As one implementation of Codd's theory, it is not without its flaws (Date, 1989b). But these flaws are more in detail than in conception and they only serve to emphasize the tremendous scope of the language and how this could be extended by intelligent enhancements to the SQL syntax.

Widely used in the business world, it is my belief that the language is so well suited to the seemingly overwhelming complexity of natural resource data, that it will eventually find some of its most important uses in that arena.

The great strength of the relational strategy is that it allows for the design of a database to be completely uncoupled from the questions that it is intended to answer. On the assumption that the proper data is being gathered, and in the proper volume, then one need not worry about how to structure the data so as to answer these questions. In relational theory, the data is arranged, rather, so that it is properly "normalized" (Lien, 1985). Normalization insures that no row in any database table will have multiple fields with precisely the same definition. If this is done correctly, it should in theory be possible answer any well-structured query one might want to ask, whatever it might be.

Normalization requires, in turn that fundamental "entities" be identified. This has led to the development of the entity-relationship model as one technique for identifying those "objects" which the database user would like to manage via individual database tables (Chen, 1985). These tables provide the primary building blocks of the database. Relationships between entities are developed into associations which form the second tier of tables. Unique sets of identifiers taken from each of the primary tables are linked through such an association providing a pathway for querying the database. In this way fields with the same definition represent a separate entity to be managed. A table is created for the new entity, and a new association defined as the link between the two entities. This association also resides in a separate table.

It is outside the scope of this paper to detail all of the entities which will be managed as part of the Starkey project, and the relationships between them. Included in this list, however, are telemetry observations, climate observations, collars, information on elk, deer, and cattle, the relationship between collars and animals, forage samples, animal introductions and removals from the range, and much more.

Telemetry data sets are periodically retrieved on tape from Starkey to the Forestry and Range Sciences Laboratory in La Grande, Oregon, about 28 miles from the Experimental Forest and Range. There they are transferred via another MS-DOS PC to a Data General MV15000 series computer, and bulk-loaded into a relational database management system purchased from *Oracle Corporation*[2]. This relational system provides us with the means of querying the data quite generally.

As useful as is the organizational framework provided by these ideas, relational theory as implemented through the SQL language takes these ideas to a new level. Views are virtual tables, whose specifications can be tailored to the needs of the moment. Once defined, the same language elements which are used to query tables can be used to select from the view as if it were a real table (thus the name virtual table). The tremendous control which views give developers over the very conception users have of what is in the database cannot be overemphasized. Simply stated in the language of the relational database world, a properly designed series of views and tables can form the basis for a dynamic entity virtualization schema.

It is precisely this conception which is at the heart of the strategy we have implemented to allow for the flexible creation of maps for the Starkey project. These ideas have, surprisingly, hardly been explored. Perhaps this is due to the relatively recent development of relational database concepts. Whatever the reason, the fact is that they provide a rich framework for mirroring the dynamic nature of natural resources data.

## The role of GIS

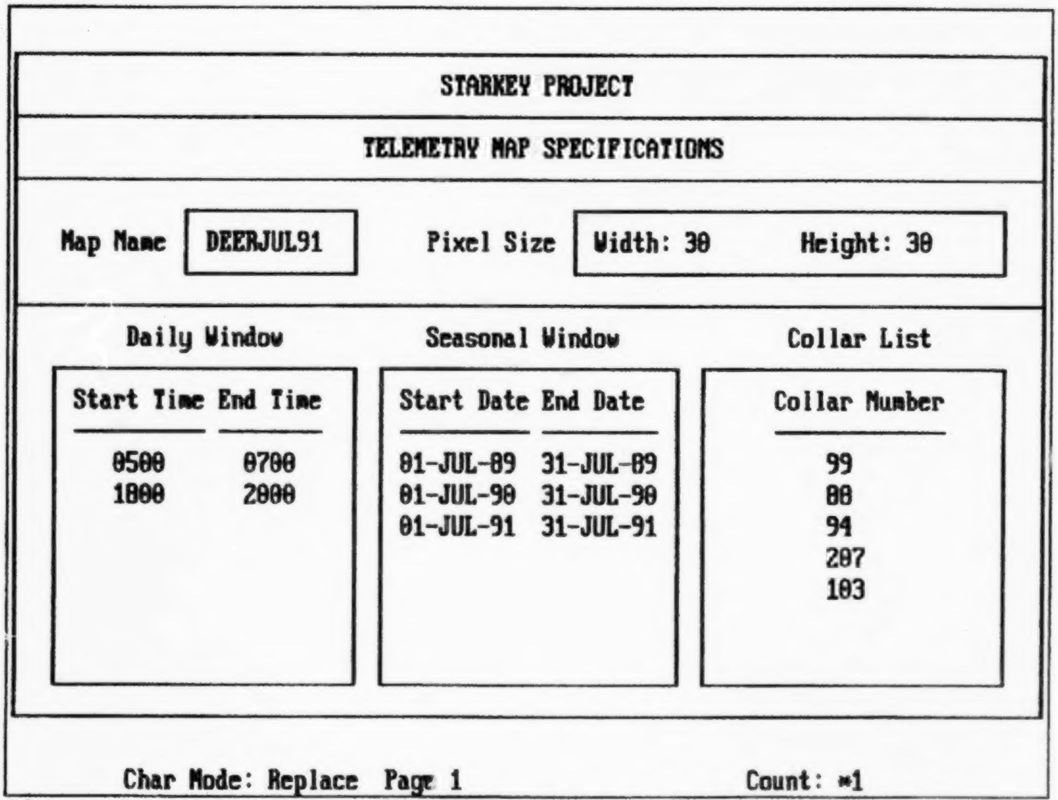
The U. S. Forest Service is currently in the process of issuing a Request for Proposals for GIS hardware and software. In the interim, those needing GIS services are granted permission on a case by case basis. Our group asked for and received permission to use the public domain MOSS system as part of our study.

Since we may eventually migrate the study to a different GIS system than MOSS, care had to be taken in planning the interface between our telemetry database and the GIS system. The intensive querying of the telemetry observations are performed by the DBMS, while the GIS performs additional classification of the resulting aggregated values, as well as the overlaying of the ensuing maps with those for slope, aspect, vegetation, topography, stream buffers, road buffers, and other subjects.

## The map machine

The primary cooperator to the U. S. Forest Service on the Starkey Project is the Oregon Department of Fish and Wildlife. Close cooperation with this agency is essential to meet the needs of project scientists and technicians. This cooperation resulted, very early in the project, in an informal requirements document. The ODF&W wildlife biologists outlined their needs in concrete terms, such as the identification of movements of various classes of animals during various periods of time.

The desired animal classes were invariably some combination of species, age, and sex, while the time periods were closely related to diurnal and seasonal cycles of animal activity and plant growth. It was also pointed out that these needs might eventually broaden to include the ability to observe the movements of additional as yet unspecified animal classes during certain undefined time periods.



STARKEY PROJECT

TELEMETRY MAP SPECIFICATIONS

Map Name: DEERJUL91

Pixel Size: Width: 30 Height: 30

Daily Window

| Start Time | End Time |
|---|---|
| 0500 | 0700 |
| 1000 | 2000 |

Seasonal Window

| Start Date | End Date |
|---|---|
| 01-JUL-89 | 31-JUL-89 |
| 01-JUL-90 | 31-JUL-90 |
| 01-JUL-91 | 31-JUL-91 |

Collar List

| Collar Number |
|---|
| 99 |
| 00 |
| 94 |
| 207 |
| 103 |

Char Mode: Replace Page 1 Count: *1

**Figure 1:** Screen entry form for telemetry map specifications

**Table I:** User definable elements for Starkey telemetry map specifications

| | |
|---|---|
| Map Name | Used to name the map and for metadata |
| Pixel Size | Aggregating resolution for resulting map |
| Daily Window | List of diurnal cycles of interest |
| Seasonal Window | List of seasonal cycles of interest |
| Collar List | List of collars worn by animals of interest |

After some thought, we were led to the conclusion that any scheme for extracting information from the telemetry table had to be quite general in order to meet these open-ended needs. The explicit solution is both simple looking and simple to use. The form into which users enter their requests is shown in Figure 1. The available "windows" into which information may be entered are listed in Table I.

The daily and seasonal nature of the time cycles of interest to the project biologists are reflected in the design of this request form. In addition, an arbitrary list of animal collars can also be targeted for observation in a third window.

In order to allow ourselves a good fit with our other data sets, the pixel resolution is also user-definable. This insures that the maps generated by the system will match the resolution of the currently available Landsat data. Since we are currently in the process of determining the what confidence interval to associate with the Loran data, this gives us the flexibility to adjust our frame of reference to this interval in both dimensions.

Each of these user definable elements is deposited into a series of small tables which have been emptied prior to the call to this form. These map specifications then interact with the data set through a view which is shown in. Here the flexibility, conciseness and power of SQL are revealed in detail. This single view, from which all maps are created, simultaneously converts time values from Greenwich Mean Time (in which all our telemetry observations are taken) to Pacific Standard Time and extracts both time and date information, while checking to see that all of the specified criteria are met.

Examining this view in more detail reveals some of the salient features of the language. The **FROM** clause indicates which tables are being used. The **WHERE** clause specifies the conditions on which the selection of elements is predicated.

The **SELECT** clause indicates which elements will be returned. This is the heart of the query. A proper grasp of the language elements at this point starts to reveal the elegance of the underlying conceptions. It is very important to realize that *the query will operate on all elements of the underlying tables which meet the specified criteria*. In other words, our basic unit of data is the table, and we always are operating on all elements of the table implicitly.

In this **SELECT** statement, a discrete function rounds off a calculated value for the pixel row and pixel column, which has the effect of mapping all observations which were located anywhere on a given pixel to the row and column number for that pixel[3].

The **GROUP BY** clause when used with the COUNT(*) function insures that each unique combination of row and column will act as an accumulation point for the aggregated count of all rows which return the same ROW_NUM and COL_NUM. We call this value VISITS, and we're done. ROW_NUM, COL_NUM, and VISITS are now available for selection from the virtual table (view) called TLM_VIEW.

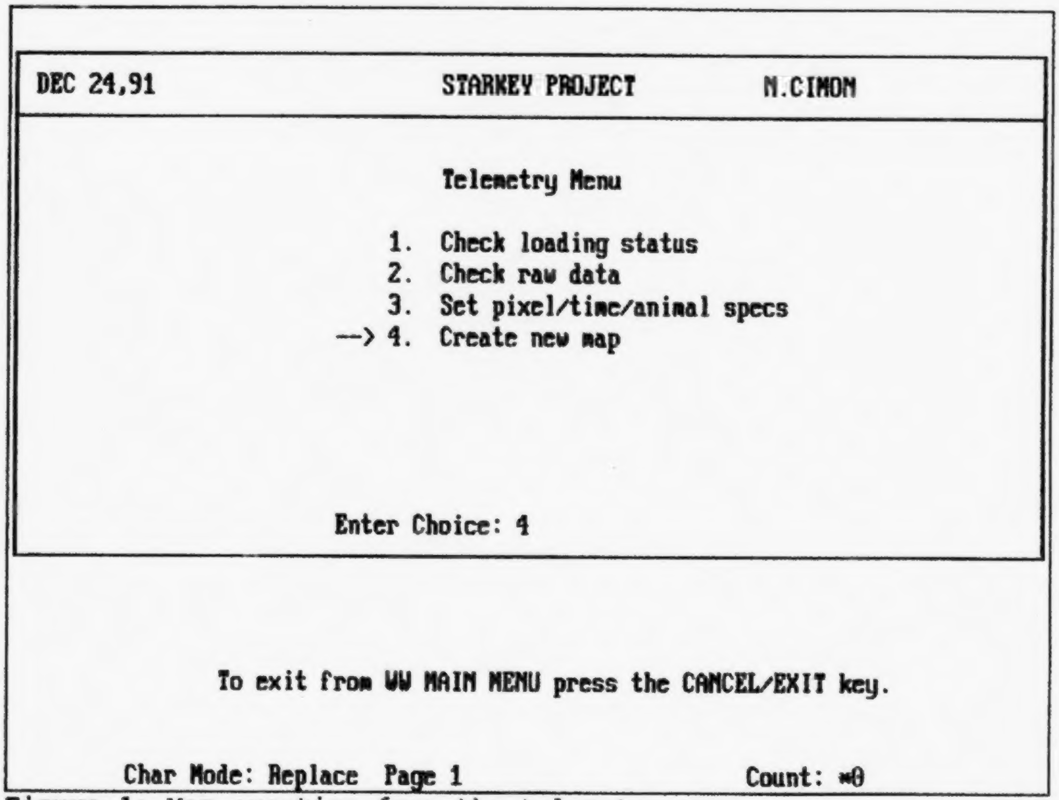
## DBMS-GIS interface

When selected from, this view reports out the row and column for each pixel which was visited, as well as the number of times it was visited within the specified criteria. The data is saved out of the *Oracle* system as an ASCII text file.

A general map format, IDIMS[4], was chosen as the vehicle for import into MOSS, and a one page Fortran program written to perform the conversion.

All of this is invisible to the user, since she or he creates a map at the same menu from which the map specifications form was originally called (). Underlying this menu selection, calls are made to the operating system to start a batch process for selecting the map data from the view, and formatting it for import into the GIS.

After successful creation of a map and prior to deletion from the specification tables these tables are joined together and inserted into a separate table which permanently records the map name and the criteria used



```
DEC 24,91                          STARKEY PROJECT          N.CIMON

                                   Telemetry Menu

                               1.  Check loading status
                               2.  Check raw data
                               3.  Set pixel/time/animal specs
                           --> 4.  Create new map



                         Enter Choice: 4



              To exit from WW MAIN MENU press the CANCEL/EXIT key.

         Char Mode: Replace  Page 1                  Count: *0
```

Figure 1: Map creation from the telemetry menu

to create it. This metadata is essential to the proper tracking of the maps. Given the broad nature of the criteria used to create maps, no map naming convention, or single line subject description could possibly be adequate. Additionally, it is in the nature of the modern DBMS to use its facilities at all levels for data and metadata management (most use their facilities to manage their own system tables).

The data sets and maps created by this system will be used to test the ways in which the different groups of animals interact, and to determine these changes visually. One interesting feature is the use of a very inexpensive micro-computer[5] to produce animations from sequences of maps. An excellent public domain Tektronix terminal emulation program[6] allows for the capture and downloading of sequences of maps created in the fashion described above. For instance, a 26 week sequence of observations taken around sunrise for elk could be animated with a fade from one image to the next, showing the changes in habitat use which occurred over this time. This process has been successfully tested on a small scale and will be used extensively.

Planned enhancements to the menu system include notification of the user via electronic mail when the newly created map is ready for import, eventual automation of the import process once a GIS system is chosen, use of the animal-collar relation to classify the animals by sex and age, and the eventual use of a floating time for sunrise and sunset.

## Discussion

It cannot be emphasized too strongly that the ability to query a centralized data repository with a single selection statement in order to create very general maps is revolutionary. The simple syntax of the statement belies the complexity of the objects being managed. These are collections of telemetry observations which represent the interplay of specific groups of animals with carefully delineated diurnal and seasonal time cycles. What is more, these sets of telemetry observations are simultaneously remapped onto the study area at any desired pixel resolution.

```
/**
*  VIEW TLM_VIEW
**
*  View to select use levels (VISITS) by pixel size (TLM_VIEW_PIX:
*  HEIGHT, WIDTH) for both daily (TLM_VIEW_TIME BEGIN, STOP) and
*  seasonal (TLM_VIEW_DATE: BEGIN, STOP) windows, and for a
*  selected list (TLM_VIEW_COLLAR) of animals
*   Referenced tables:
*       TLM             (BASIC TELEMETRY INFORMATION)
*       TLM_VIEW_PIX    (ANALYSIS PIXEL SIZE)
*       TLM_VIEW_DATE   (THE DATES FOR THE ANALYSIS)
*       TLM_VIEW_TIME   (THE TIMES FOR THE ANALYSIS)
*       TLM_VIEW_COLLAR (THE COLLARS USED IN THE ANALYSIS)
*
*  Notes:
*       - ROUND function is used to discretize the TLM UTM_X, UTM_Y
*         values and produce the column and row numbers (COL_NUM,
*         ROW_NUM) for the pixels.  These discrete pixels then become
*         the "buckets" for collecting VISITS, via the GROUP BY clause.
*       - TO_CHAR function is used to extract date and time information
*         from the OBS_TIME field in order to obtain values that
*         can be tested against the specified daily and seasonal
*         windows.
*       - NEW_TIME function is used to convert the OBS_TIME column from
*         Greenwich Mean Time (GMT) to Pacific Standard Time (PST).
**/
CREATE VIEW TLM_VIEW AS
  SELECT ROUND((UTM_X-373000.0)/WIDTH+.5)   COL_NUM,
         ROUND((5020000.0-UTM_Y)/HEIGHT+.5) ROW_NUM,
         COUNT(*)                           VISITS
  FROM   TLM,
         TLM_VIEW_PIX,
         TLM_VIEW_TIME,
         TLM_VIEW_DATE,
         TLM_VIEW_COLLAR
  WHERE  TO_CHAR(NEW_TIME(OBS_TIME,'GMT','PST'), 'HH24MI')
          BETWEEN TLM_VIEW_TIME.BEGIN AND TLM_VIEW_TIME.STOP
  AND    TO_CHAR(NEW_TIME(OBS_TIME,'GMT','PST'), 'DD-MON-YY')
          BETWEEN TLM_VIEW_DATE.BEGIN AND TLM_VIEW_DATE.STOP
  AND   TLM.COLLAR = TLM_VIEW_COLLAR.COLLAR
  GROUP BY ROUND((UTM_X-373000.0)/WIDTH+.5), ROUND((5020000.0-UTM_Y)/HEIGHT+.5)
/
```

**Box 1:** View definition for creating maps from specifications

The older procedurally based programming paradigm, with its record by record selection and manipulation of data, would have required hundreds, perhaps thousands of lines of code to accomplish the same feat, without providing the forms environment, and the well-defined structures for storing and retrieving data and metadata in a flexible manner.

The SQL language is deceptively simple looking. This superficial simplicity masks great power. The ability to virtualize the entity being managed allows for open-ended objectification and the idea of maps on demand.

In a field which is rife with extravagant claims, the relational model for database management has held up remarkably well. Codd's original formulation of a logical framework for large data sets has travelled the road from abstraction to implementation and now offers one of the very best solutions for the manipulation of the complex data sets so common in the field of natural resources. While SQL, as one implementation of the relational model has its flaws, the strong mathematical foundation provides a framework for future enhancements to the language.

## Bibliography

Bracken, P. A. [1983], Remote sensing software systems: IDIMS software system, in *Manual of Remote Sensing* 2nd edition, Vol. 2, pp 824-826. American Society of Photogrammetry, Falls Church, Virginia.

Chen, P. [1985], Database design based on entity and relationship, in *Principles of Database Design*, Vol 1. Logical organizations, chapter 6. Prentice-Hall: Englewood Cliffs, New Jersey.

Codd, E. F. [1970], A relational model of data for large shared banks. Communications of the CACM, 13 (6), pp 377-387.

Dana, P. H., Fowler, W., and Hindman, D., [1991], Automated animal-tracking system: tracking elk with retransmitted Loran-C.

Date, C. J. [1989a], Why SQL is important: background, chapter 1, section 1, in *A Guide to The SQL Standard* 2nd edition. Addison-Wesley: Reading, Massachusetts.

Date, C. J. [1989b], An annotated SQL critique, Appendix F in *A Guide to The SQL Standard* 2nd edition. Addison-Wesley: Reading, Massachusetts.

Johnson, B. K., Noyes, J. H., Thomas, J. W., and Bryant, L. D., [1991], Overview of the Starkey project: current measures of Elk vulnerability, in *Proceedings of Elk Vulnerability Symposium*, in press.

Lien, Y. E. [1985], Relational database design, in *Principles of Database Design*, Vol 1. Logical organizations, chapter 6. Prentice-Hall: Englewood Cliffs, New Jersey.

Mendelson, E. [1964], *Introduction to Mathematical Logic*. D. Van Nostrand Company, New York, New York.

SLAS [1990], *A VLT User's Manual*, SLAS Report 370, August 1990, Stanford Linear Accelerator Center, Stanford University, Stanford, California.

[1] *Specifically, the predicate calculus and the theory of relations form the foundation for the relational model. These are deep subjects which touch on how we quantify and order the objects of the world. They can be explored more completely in books such as Mendelson, 1964.*

[2] *The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the U.S. Department of Agriculture of any product or service to the exclusion of others that may be suitable.*

[3] *The UTM coordinate point (373000.0, 502000.0) represents the northwest corner of the northwestern most pixel for all currently registered GIS maps.*

[4] *The format used by the Interactive Digital Image Manipulation System. See Bracken, 1983.*

[5] *An Amiga 2000 from Commodore Business Machines.*

[6] *VLT, available from the Stanford Linear Accelerator Center (SLAC, 1990).*

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Integrating spatial models of recreation, wildlife, and timber resources

**Lloyd P. Queen**

Department of Forest Resources
1530 North Cleveland Avenue
University of Minnesota
St. Paul, Minnesota 55108

## Abstract

*Integrated resource management (IRM) involves the simultaneous consideration of multiple management perspectives. One goal of IRM is to ensure that all resource perspectives are considered at the same level, and that interdependencies between those (sometimes different) viewpoints or perspectives be explicitly considered. GIS provides a vehicle for integrating diverse resource perspectives to aid in resource management decisions. The research presented here focuses on developing GIS-based data and process models for integration of recreation, wildlife, and timber resource management strategies. Recreation concerns are addressed through application of the Recreation Opportunity Spectrum (ROS); wildlife issues center on the use of Habitat Suitability Index (HSI) models for several target species; and timber management is primarily centered on scheduling of timber harvests to maximize economic value. Each of these three resource models (ROS, HSI, and harvesting) were implemented for a multi-ownership forested area in northern Minnesota. Structured needs assessments conducted prior to model implementation resulted in a robust GIS database design that was capable of supporting all three fields of activity. Process models in network format were also developed for each model individually, then later combined into a single, integrated model that added explicit spatial context to each of the three individual models. Lastly, data and other functional interdependencies were examined in light of the impacts of forest roads on model results.*

## Introduction

Today, resource managers have begun to embrace geography as a mechanism that can be used to facilitate integration of multiple resource perspectives. Beyond recognition of shared geography as being common to foresters, recreation planners, wildlife specialists, hydrologists, and others who work in a specific area; the advent of Geographic Information Systems (GIS) technology has increased managers collective abilities to (essentially) simultaneously consider multiple resource perspectives.

Over the past half-decade, several GIS-based efforts at completing integrated resource assessment and management projects have appeared in the literature (see for example, McRae and Cleaves, 1986; Klock et al., 1986; Lam et al., 1987; Reisinger and Davis, 1987; Tomlin and Johnston, 1988; Brand and Becker, 1991; Riekena, 1991; and Leggat and Buckley, 1991). As a result of this considerable effort, several key issues have repeatedly been raised as keys to successful integration. This spectrum of issues includes personnel, organizational, as well as technical matters. However, perhaps the central issue in implementation of integrated projects and programs is the development and use of a combined spatial and attribute data base. This paper adopts a data-centered approach to integration, and will discuss the use of geographic data for multiple resource integration.

Land data share a number of features across virtually all disciplines. Data tend to voluminous, they tend to have a highly variable information content, they are not shared vertically or horizontally within and between users and organizations, and they tend to be unique to user, place, and time. As a result, the single resource databases so common today tend to be incompatible; leading to analyses which can be non-specific and poorly integrated. Data administrators also then, tend to expend considerable time on data exchange and input standards, so that when users call for multiple resource data, their data needs can be met.

As the integration paradigm has developed, numerous definitions of what constitutes integration have evolved. Most often, integration is viewed as the simultaneous consideration of multiple resource perspectives. Taken another way, integration can be thought of as the sharing of data from one technology or system (user) to another. In each case, those persons sharing data require that full and complete recovery of their original data is assured. Indeed, although this paper focuses on data integration, users and their support technology also are necessarily subject to the need for integration. In fact, the work reported here integrates user needs and concerns one level beyond data — to the actual processing of spatial data via models implemented in a GIS environment.

Several techniques which we adopted to facilitate data integration include:

| Techniques for Integration |
|---|
| Lifecycle concept |
| Explicit model building |
| Imposition of standards |
| Use of neutral variables |
| Role reversal |
| Mutual geography |

In every case, successful integration must be predicated on a comprehensive development plan. The lifecycle concept, popularized by Somers (1989) is perhaps the most widely recognized strategic approach for system and program development. A lifecycle approach considers the long-term issues associated with programmatic efforts (planning, design, implementation, and maintenance) and thus can help to safeguard "integrability" over time. Significantly, the major aspect of planning is a structured user requirements analysis. These analysis techniques are applied to extract user data and information needs, and would normally precede all but the smallest GIS projects (see McRae and Cleaves, 1986). Essentially, structured planning efforts abstract what data users require, how users intend to manipulate data (model building), and what information products are needed by end users. In a multi-participant project a clear understanding of user needs is prerequisite to successful integration. Structured needs assessment tools explicitly examine the data and model interdependencies that are the basis for integration in a GIS database environment.

As a result of extracting user data needs and modeling plans, data standards may be specified for all project participants. Standards are used to assure some minimum compatibility between data sets. Typically, data standards preserve the ability to transfer data between systems (users). However, specification of data standards may also encompass information content and accuracy, as well as data tracking procedures such as dictionaries and catalogs that inform all users as to the nature of the data contained in the database.

Each of the three integration vehicles discussed above might be applied even to single-resource projects, in anticipation of integration which might occur at a later time. The final three techniques mentioned here are proactive approaches actually applied during completion of integrated efforts. Our project made use of forest roads as a neutral environmental variable. That variable affects all resource perspectives involved in the project. While no single participant has ownership of the neutral variable "roads," roads do affect each resource perspective (timber, wildlife, and recreation). And so, by requiring each participant to consider the impact of this variable, some mutuality is assured. Users thus react to a common goal (the impact of forest roads), which leads directly to impacts presented by each other resource area's activities.

Role reversal simply involves allowing each resource specialty to be both proactive and reactive with respect to all other perspectives. True integration must allow each participant to operate at an equal level. In forestry applications such as the one reported here, it is not uncommon for timber harvesting to provide the catalyst for change; to which all other managers are asked to respond. Role reversal allows recreation planners, for example, to set conditions that impact the domain of possible actions that can be undertaken by either the wildlife or the timber manager.

Finally, while GIS does provide unparalleled opportunities for integration at the data level, users must look beyond shared or mutual geography as the sole integrating parameter. By applying the other techniques mentioned above, issues that prevent integration (such as incompatible scales, accuracy levels, and information content) can be overcome. Although we can envision that GIS provides a framework for a database that could be supplemented with other data to support all resource management, shared geography alone is inadequate to the task of integration (see Klock et al., 1986).

The remainder of this paper discusses, in turn, project goals, the efforts of each of the three individual resource perspectives, GIS development, and a summary statement on the design of integration approaches.

## Project goals

In 1989 the Legislative Commission on Minnesota Resources funded a multi-participant project aimed at investigating the impacts of forest roads on timber, tourism, wildlife, and aesthetic resources in northern Minnesota (for a full project description see Rose et al., 1991). A small group of researchers from each of the resource disciplines was established as a project management team. A critical member of that team was the GIS facilitator; who in fact played the role of project integrator. The following goals were established by these project managers for a 19,000 hectare (two township) test area located in Itasca County, Minnesota:

1. Establish a multi-participant, multi-resource team to assess road impacts and give each member both proactive and reactive roles in that assessment;
2. Use a structured design process to ensure parallel data and model development and integration;
3. Use GIS, database management systems (DBMS), and modeling to describe existing interactions between disciplines; and
4. Use the GIS and DBMS to assist in the production of an integrated, seamless database and to provide a comprehensive analysis capability.

Separate models were eventually created to model each of the three resource perspectives; recreation, wildlife, and timber management. Essentially, the project had three phases; 1. individual resource model development, 2. data development, and 3. project-wide integration. Individually, the timber management model used an area control strategy to produce harvest schedules over time. The wildlife management model was centered on evaluation of habitat for chosen species and the overlap between wildlife habitat and human activities. Human uses of forests for recreation was modeled using experiential recreation opportunities, physical recreation opportunities, transportation network use, and the aesthetic impacts of forest roads.

Each member of the project was subject to data and model development standards imposed by a project facilitator. Forest roads impact each of the project components, and thus were established conceptually as the center of the project as a neutral environmental variable. The key point in the overall design was the connectivity between project components. The design is predicated on interaction and shared relationships between data, management goals, location, and modeling approach.

Perhaps the single most unique aspect of the overall project approach was the design goal of integrating each individually developed project component (timber, wildlife, recreation) into an overall framework that explicitly modeled interrelationships and linkages between management approaches. This technique is in contrast to implicitly considering relationships between timber, recreation, and wildlife management based solely on the fact that each assessment or management person happens to be working in the same geographic area. Also, by designing an integrated project from the outset, each person could benefit from data or model development performed by other team members.

## Timber management

Timber management primarily consists of the operations performed on a site for the production of timber resources. These include planting, fertilization, thinning, harvest, skidding, and transportation. With the emergence of aspen as the primary wood producing species in Minnesota, our initial assessment was centered on that one species. In order to access the aspen resource, a simple area control strategy was applied to aspen harvest.

Area control is based on the assumption that it is desirable to achieve a balance of age classes over time. The amount of acreage cut in each period is simply the total acreage divided by the number of time periods. For example, if there are 6000 hectares of aspen and we wish to achieve a balanced age class distribution in 60 years using 10-year periods, we would harvest 1000 hectares per period (100 ha per year). Typically, a hierarchical strategy is used to determine the order in which stands are cut; the usual strategy is one of cutting the oldest stands first.

In our application, aspen stands were assembled into a vector of age-based records together with acreages for each stand. The allowable cut was derived as the total area divided by rotation age. For a stand-based inventory such as that available in this project, the output of area control modeling effort was a schedule of harvests occurring in each harvest period. Thus the spatial distribution of different aged aspen stands was manipulated by area control strategies.

## Wildlife management

Due to the size (two townships) of the study area there were several constraints on wildlife assessments. Because of the small size of the area, an analysis of species such as wolves, which range over much larger areas, was precluded. Secondly, we attempted to develop models for species which are relatively well-understood. Since the intent of the project was to consider impacts of roads and other resource perspectives on wildlife, we only wanted to demonstrate issues of concern to wildlife managers. Lastly, we were constrained by the available base-line forest data. Because we were constrained to stand data collected as part of a commercial forest inventory, models had to be developed which conformed to these data. It is important to note that our intent was to develop techniques to apply wildlife models and to demonstrate ways of obtaining information which may assist in wildlife planning efforts.

The base model for analysis of habitat protection and development was the Habitat Suitability (HSI) Model. Rose (1991) gives a review of the evolution of HSI models. These models are applicable in that they assess existing habitat quality for a given species, and will adjust for different prevailing conditions (e.g., timber management practices).

HSIs are based on minimum areas of applicability and on the ability of that area to meet the life requirements of a species. For our purposes, an HSI model was developed for ruffed grouse in areas where aspen provides for the life requirements of the species. Research has shown that different age classes of aspen meet the food and cover requirements of ruffed grouse, so the model was developed to assess the year-around quality of aspen habitat for the grouse. The minimum habitat area for our HSI model is approximately 5 hectares. Within that minimum area, optimal grouse habitat is found where there is an equal distribution of the different age classes of aspen. The availability of all age classes of aspen (0-12 years; 13-25 years; 26-38 years; and 38+ years) in close proximity will provide for brood cover, winter, spring and fall cover, nesting, and food requirements, respectively.

A simple algebraic expression was derived to assess the spatial variability and age class distribution of aspen for the minimum habitat area. The model is:

$$\mathrm{HSI} = (\mathrm{SIV}_1 + \mathrm{SIV}_2 + \mathrm{SIV}_3 + \mathrm{SIV}_4)/4$$

where the $\mathrm{SIV}_i$ terms are the percent of aspen in each age class (i= 1,2,3, or 4) for the analysis area. This model was applied to the rasterized aspen stand coverages using convolution filters which were sized to the appropriate minimum habitat area (approximately 5 hectares).

Basically, HSI values are a function of the spatial and temporal distribution of aspen. Aspen cover type maps exist at some initial condition (current forest inventory) and will change based on management and harvesting scenarios. By the same token, HSI scores can be achieved for an area by adjusting the age class distribution via various management activities.

## Recreation management

Roads are an integral part of the recreation opportunities that exist in any given area. The existence or lack of roads dictates the type of recreational experiences available to the recreationist. Areas lacking roads provide semi-primitive and primitive experiences. Areas with roads provide access to recreation facilities and are often a recreation facility unto themselves.

For this effort, four categories of road impacts were developed:

1. Experiential recreation opportunities (roads alter the natural landscape and thus the types and nature of recreation opportunities);
2. Physical recreation opportunities (access to facilities is necessary to the recreating public);
3. Transportation network (considers driving for pleasure as a recreation opportunity); and
4. Aesthetics (those persons using the transportation network are provided with a recreational viewing experience).

For actual management convenience, possible mixes of settings, activities, and probable experience opportunities such as those described above were arranged along a previously developed spectrum or continuum (see Rose et al., 1991). This continuum is called the Recreation Opportunity Spectrum (ROS), and is divided into six classes, each defined in terms of it's combination of activity, setting, and experience opportunity:

| Recreation opportunity spectrum classes |
|---|
| primitive |
| Semi-primitive non-motorized |
| Semi-primitive motorized |
| Roaded |
| Rural |
| Urban |

Using ROS guidelines distances from different types of roads are used to define a base map of the six ROS classes. Minimum size requirements for the primitive and semi-primitive classes are used in a reclassification scheme to produce a final ROS map. Any change in the road network will produce impacts that are modeled by repeating ROS definition on the revised roads map. Outputs from the ROS modeling effort are maps showing the locations, shapes, and sizes of the various ROS classes found in the study area. Another product is simply the dynamic ability to define the impact of forest road changes on a management area.

## GIS development

In terms of technical support, three separate GIS software packages were required to meet all stated user needs. Indeed, multi-resource projects should anticipate user needs that cannot be met by any one system or software package. In this type of multi-system environment, data standards are critical to making sure that any one user operation or system-based process does not adversely affect the project-wide utility of a data set. In terms of data tasks, a vector-based GIS was used for development, and both raster and quadtree systems were used for analysis and modeling. Sharing of data between systems was made possible through the use of data format translators. Because users were spread between multiple, remote, non-networked sites, data sharing and maintenance protocols were necessarily established and enforced by the GIS facilitator.

Conceptually, most users envision a GIS as a stacked set of map "layers," where each layer corresponds to a map theme. While this model is useful as a user view, it is impractical in terms of data management. The project thus made use of a relational database management system to handle the multiple attributes tagged to each mapped entity. Forest stands, for example, are characterized by a large number of inventoried variables. Each of these stand attributes was linked to the stand map using the geographically relational database. All maps and attribute files were compiled and stored using a single scale, coordinate, projection, and classification system.

GIS procedures supported information extraction for each user. For some team members, the types of geographic questions were very basic. The timber management people required information, for example, on the distance of stands to roads and mill locations. While this is simple information in a static sense, the GIS automated the extraction of this type of data, and would allow the user to quickly regenerate this data should road or distance conditions change (in reality or in a simulation). In other cases, the spatial variability of mapped data was the fundamental basis for a team's approach. Calculation of a habitat suitability index for ruffed grouse, for example, required assessment of the distribution of the different age classes of aspen. This type of operation is too impractical to perform manually, and could not effectively be performed without the windowing capability afforded by raster GISs.

## Integration

The integration design was predicated on a structured process adapted from systems analysis procedures. This strategy began with a User Requirements Analysis (URA), which queried each user as to their data needs and modeling approach. Essentially, the URA inventoried the goals of each individual user. The URA was then used to establish matrices of data structured by theme, attribute, and user needs. By maintaining a centralized database, equal access for all participants to completely seamless, standardized data and coverages was secured. Also, by explicitly generating these data matrices, apparent linkages between users were shown.

Another benefit derived from the URA was knowledge of data manipulation and analysis needs. Project managers were able to anticipate team needs and could thus ensure that individual model developments were occurring along parallel tracks. That is, each user consulted with project leaders as to implementation options for data analysis. Models were selected which used existing data and analysis procedures available in the GIS database. Essentially, GIS provided a spatial framework or template, within which each team member was required to work.

One difficulty often encountered in multi-resource assessment or management projects is domination by a specific user group or team. Often times this influence or control stems from one group's (resource perspective) having better data, more robust models, or more clearly defined objectives. This situation thus places other project participants in more of a reactive role; where the dominant group's goals are, by default, given priority. An additional design goal of this project was to place each user on equal footing; making sure that linkages between all persons (in terms of data as well as modeling approach and management objectives) were explicitly defined. Initially, existing conditions maps were created for each team that characterized the landscape within the context of their broad management goals. Simulations were then conducted within and between each resource context, and thus interrelationships between perspectives were explored.

Lastly, all team members required data inputs which were derived from the outputs of other team simulations. Output products were generated according to system specifications set by the project management team, and the rigid structuring of data within the GIS database ensured a seamless sharing of data. Where appropriate, participants communicated design or data formatting goals to each other in order to minimize data reformatting and logical conflict problems.

In summary, integration was designed into the project. Participants were guided along parallel tracks in designing and in implementing models. A data-centered approach was initiated and maintained, and data were structured according to shared geographic rules established within the GIS. Linkages between team members were sought in proactive and reactive modes, and linkages were defined via design diagrams which documented team interrelationships (see Figure 1).

The most basic linkage between persons involved in the project came from shared geography and shared resource data. The data needs of each team member, as defined via the User Requirements Analysis, were quite similar. Primary resource assessment data provided by the U.S. Forest Service, the Minnesota Department of Natural Resources, and the Itasca County, Minnesota Land Department were all map-based. Thus, GIS procedures were used to manage and to provide basic spatial processing capability for all data.

An important predicate for GIS use is a structured plan for system development, implementation, use, and maintenance. Again, the User Requirements Analysis

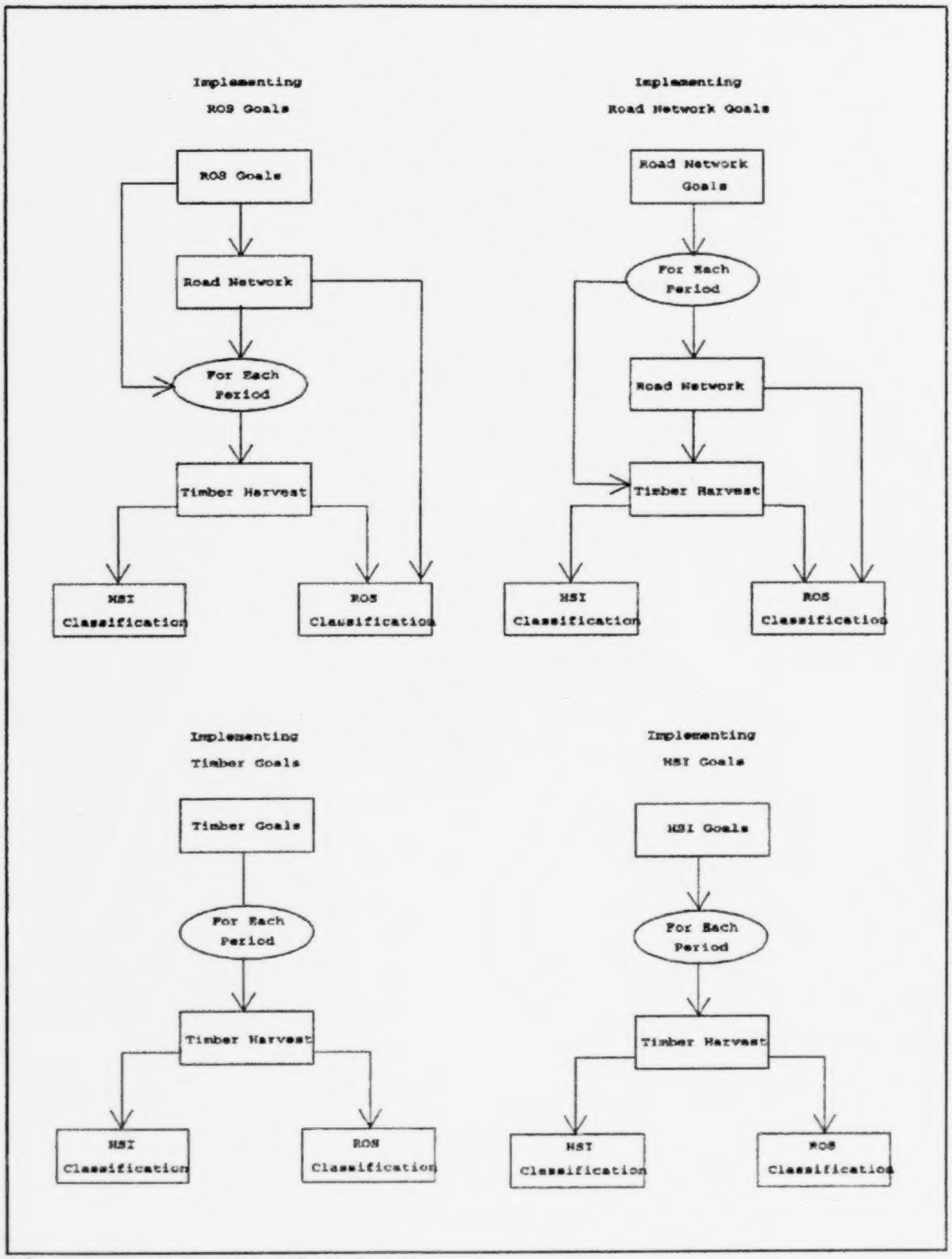

Implementing
ROS Goals
ROS Goals
Road Network
For Each Period
Timber Harvest
HSI Classification
ROS Classification
Implementing
Road Network Goals
Road Network Goals
For Each Period
Road Network
Timber Harvest
HSI Classification
ROS Classification
Implementing
Timber Goals
Timber Goals
For Each Period
Timber Harvest
HSI Classification
ROS Classification
Implementing
HSI Goals
HSI Goals
For Each Period
Timber Harvest
HSI Classification
ROS Classification

Figure 1 Impacts Flowchart

provided the basic design goals for the GIS. Data could then be developed and integrated into the GIS based on a set of common rules. These rules and standards were defined and maintained in a project-wide data dictionary.

## Conclusions

GIS provided the basic vehicle for integrating the various project approaches. Data proved to be a fundamental link between team members, and the rigid structuring required by the GIS secured the ability of any project participant to share and access data. Data development was a very costly and time-consuming task. The geographic base files generated during this project represent a critical deliverable that could surely be used by other researchers to address additional integrated resource management questions. Also, the process used to develop this GIS approach will be invaluable to resource agencies in Minnesota as they adopt GIS techniques.

GISs are more than computer hardware or software and are more than a database; they are in themselves integrated systems for managing and analyzing spatial or mapped data. Because mapped data are fundamental to all resource assessment and management, virtually all resource agencies are getting started with GIS. This project broke new ground in Minnesota with respect to developing data in support of integrated resource management. Insofar as the system was designed to support specific resource assessments, the GIS developed for this project could be termed a Spatial Decision Support System. Certain GIS functions were optimized for specific user requirements, and the GIS played a technical support role to all of the team members. A next logical step would be to streamline the system and to customize the user interface in order to provide access to the power of a GIS without requiring resource managers to incur all of the technical overhead associated with generic GIS approaches. These types of Decision Support Systems represent a significant potential for further research.

In conclusion, there are several key benefits to a GIS approach to integrated resource management. GISs can be data-focused, they remove computational drudgery, they enforce algorithmic thinking and encourage holistic thinking, they are interactive, and they facilitate most of the basic principles that govern integrated resource assessment. It is clear that multiple resource assessment and management will continue to develop as a major resource paradigm. As resource professionals we should anticipate the spatial context of resource problems and of our resource data, and make every attempt to re-engineer the human aspects of GIS-based data management and analysis.

## References

Brand, D.G., and A. Becker. 1991. A Framework for Integrating Silviculture Survey with Data Analysis, Planning, and Database Update. *In GIS Applications in Natural Resources.* M. Heit and A. Shortreid, eds. GIS World, Inc. pp. 237-240.

Klock, G.O., P. Gum, and L.E. Jordan, III. 1986. Integrated Resource Management Using Multispectral Imagery and GIS: Okanogan National Forest. *Proceedings, 1986 ACSM/ASPRS Annual Convention.* pp. 46-53.

Lam, N.S., P.J. Grim, and F. Jones. 1987. Data integration in Geographic Information Systems: An Experiment. *Proceedings, 1987 ACSM/ASPRS Annual Convention,* Baltimore, MD. pp. 53-61.

Leggat, K., and D. Buckley. 1991. Implementing GIS into Alberta's Integrated Resource Planning Program. *In GIS Applications in Natural Resources.* M. Heit and A. Shortreid, eds. GIS World, Inc. pp. 251-258.

McRae, S., and D. Cleaves. 1986. Incorporating Strategic Data-Planning and Analysis Techniques in GIS Design. *ASPRS Proceedings, GIS Workshop.* pp. 76-86.

Reisinger, T.W., and C.J. Davis. 1987. Integrating Geographic Information and Decision Support Systems: A Forest Industry Application. *Proceedings, GIS '87. ASPRS and ACSM Second Annual International Conference, San Francisco, CA.* pp. 578-584.

Riekena, J.K. 1991. The CAMAS Project: Building a Computer-assisted Management and Analysis System for Natural Resource Managers. *In GIS Applications in Natural Resources.* M. Heit and A. Shortreid, eds. GIS World, Inc. pp. 241-244.

Rose, D., E. Geisler, G. Arthaud, J. Rickers, J. Vlaming, L. Queen, D. Walters, A. Ek, D. Lime, and P. Jordan. 1991. Impacts of Present and Potential Forest Road Systems on Timber, Tourism, Wildlife, and Aesthetic Resources. *Final Project Report. Legislative Commission on Minnesota Resources.* St. Paul, MN. 148 pp.

Somers, R. 1989. Management Topics in GIS Development. *A URISA Workshop.*

Tomlin, C.D., and K.M. Johnston. 1988. An Experiment in Land-use Allocation with a GIS. *ACSM/ASPRS 1988 Annual Convention: Technical Papers.* pp. 23-34.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# A biophysical process approach to wildlife modelling in forestry

**T.W. Norton**

Research Fellow
Centre for Resource and Environmental Studies,
The Australian National University, GPO Box 4
Canberra, Australia 2601

**D.W. McKenney**

Chief, Forest Resource Economics Section
Forestry Canada - Ontario Region
Box 490
Sault Ste. Marie, Ontario, P6A 5M7, Canada

**W.R. Watt**

Program Biologist Wildlife Habitat
Northern Forest Development Group
Ontario Ministry of Natural Resources
Timmins Regional Office
60 Wilson Ave., Timmins, Ontario, P4N 2S7

## Abstract

*Wildlife considerations have become a major component of integrated forest management in many nations. Planning approaches in many forest regions of North America are developing to include "habitat supply analysis" and "habitat supply modelling". These approaches and their use in Canada are reviewed briefly. A process-based approach to wildlife modelling within a geographic information system framework is then described using a case study on the rare and endangered Long-footed Potoroo,* Potorous longipes *from the production eucalypt forests in south eastern Australia. The approach employs primary site-attribute data on climate, terrain, substrate and other factors to model the key biophysical processes that underlie and modulate many biological patterns within forests. The implications and advantages of the habitat supply and biophysical modelling approaches to integrated forestry management are considered.*

## Introduction

Increasing public concern for the environment, the conservation of biodiversity and the maintenance of ecosystem processes has led to the initiation of a range of environmental strategies by governments in North America and Australia. These strategies are aimed at the sustainable use of natural systems but most are yet to be fully developed, implemented and monitored. In Australia, for example, a national strategy for the conservation of biodiversity is being developed. At the same time, both the Canadian and Australian federal governments have agreed to be parties to an international convention on biodiversity conservation at the United Nations Conference on Environment and Development (UNCED) in Brazil in June 1992.

In terms of forests, the conservation of old growth ecosystems and wildlife are of particular public concern. The proposed logging of a range of old growth eucalypt forests in Australia has led to major political confrontations, as has the proposed exploitation of old growth stands of red and white pine in eastern Canada and Douglas fir forests and associated assemblages in the Pacific Northwest (Alaska, Washington, Oregon, British Columbia). Similarly, the conservation status of forest owls and other species reliant on old growth forest ecosystems is of concern on both continents and has led to serious conflicts over resource allocation and management.

In this paper, some aspects of the evolving nature of the integration of wildlife considerations with traditional forestry management are discussed. Current developments in North America, especially Canada, based on the concepts of habitat supply analysis (HSA) and habitat supply modelling (HSM) are reviewed briefly. Another approach to wildlife modelling within a GIS framework is then described using the rare and endangered Long-footed Potoroo, *Potorous longipes* which inhabits production eucalypt forests in south east Australia. The implications and advantages of these approaches to wildlife modelling are then discussed.

## Habitat supply analysis and modelling

Habitat supply analysis and modelling are increasingly used as an aid to integrated planning and management in forested ecosystems. In Canada, most provincial natural resource agencies have research programs developing and applying HSM to limited field situations and some are considering the implementation of HSM on a provincial scale. The development of quantitative modelling approaches to wildlife habitat management is at least partly driven by the recognition that current zoning systems and guidelines to modify timber management prescriptions fail to account for forest dynamics and hence may lead to unpredictable future forest conditions and hence unpredictable conditions for wildlife.

The general approach supporting the conceptual development of HSA/HSM was advanced in the United States in the late 1970s and extended in the 1980s with the development and testing of Habitat Evaluation Procedures and Habitat Suitability Indices (eg. Allen 1982, Verner *et al.* 1986, Allen *et al.* 1987), and the development of forest simulation models which allowed managers to forecast future forest conditions under different management scenarios (eg. Boyce 1980, Covington *et al.* 1988). Recently, managers and researchers have extended this general approach to the boreal forest ecosystems of Canada; eg. Patch (1987) in New Brunswick, Marcot *et al.* (1988) in British Columbia, Bonar *et al.* (1990) in Alberta, and Watt (1990) in Ontario. This work has seen the linking of operational forest simulation and planning models with procedures for habitat-suitability evaluation.

Two main types of models are employed for temporal predictions. The first type uses simulated changes in stand dynamics over time as the criterion for defining habitat quality (eg. Martelle 1990). Duinker *et al.* (1991) termed this approach the "dynamic habitat, static population" model. The second approach, "dynamic habitat, dynamic population" combines a habitat simulation model and a wildlife population simulation model (Duinker 1986, Duinker 1989). The latter approach is an attempt to accomodate the fact that the population dynamics of wildlife species do not necessarily respond immediately, and are not necessarily related linearly, to extrinsic changes in their habitat.

Both types of models may be employed with or without a broader spatial or landscape context. Greig *et al.* (1991) termed the former analysis "spatial projection" and the latter analysis "non-spatial projection". For example, in the most simple case of non-spatial projection a single forest stand may be modelled in isolation to the surrounding forests, using a dynamic habitat/static population model or a dynamic habitat/dynamic population model. This approach may also be employed for several contiguous or non-contiguous stands (eg. within a forest block) or the entire area of the target forest to provide an estimate of the "total" habitat available for wildlife species over specified time intervals. The key point here is that while the spatial pattern of stands within a forest may have been known from basic forest inventory, these data are not used for analytical or cartographic purposes. In contrast, spatial projection employs data on the spatial configuration of stands to map future habitat patterns, and/or to address dynamics across forest stands as they relate to target wildlife species (eg. considerations of habitat connectivity, dynamics between adjacent and/or nearby stands).

The net result of these evolving approaches to HSM is that four general categories of model can be identified:

(1) Non-spatial habitat models: models that estimate habitat supply within a defined area through time based solely on attributes derived from forest inventory which are considered important components of the target species' habitat;

(2) Spatial habitat models: as in (1) but the models have an explicit spatial component embedded in the primary analyses;

(3) Non-spatial wildlife population models: models that focus on the dynamics of wildlife populations within a defined area and, using forest inventory data and knowledge of the biology of the target species, estimate habitat supply through time; and

(4) Spatial wildlife population models: as in (3) but these models have an explicit spatial component incorporated into the primary analyses.

## Application of HSA/HSM

In brief, the development and application of HSA/HSM can be summarised generally as:

(1) the development of HSA/HSM in North America and especially in Canada has been largely driven by the needs of forest managers to account for wildlife habitat in their planning efforts. Hence, most of the developments have involved adapting existing forest planning tools (eg. FORMAN: Wang *et al.* 1987, and HSG: Moore and Lockwood 1990) and explaining wildlife habitat relationships in terms of forest stand descriptions (eg. Thomas 1979, DeGraaf and Rudis 1986);

(2) models have generally been used to predict the dynamics of a single species as opposed to assemblages of species. The target wildlife species are typically large vertebrates which are commercially-exploited (ie. hunting, trapping), although modellers in at least three Canadian provinces are developing relationships and models for some species based on their conservation status (rare, threatened, endangered, potentially susceptible to intensive forestry practices), and their perceived potential as ecosystem "indicator" species (Patch 1985, Bonar *et al.* 1990, Watt 1990). The notion of indicator species is that their presence and abundance/ fecundity within an ecosystem may be indicative of the health and viability of the ecosystem;

(3) study areas range from thousands of hectares to, at least in one case, over a million hectares (eg. Hinton Forest Management Area, Alberta, Bonar *et al.* 1990). Generally, the size of the units evaluated in HSM is adhocly related to the area of the forest under study. For example, Martelle (1990) employed a unit size of 400 ha for evaluating a 40,000 ha area (ie. 100 units for the region). In the case of the Hinton FMA, the area is divided into five working circles, each with 135 compartments based on operational planning concerns (ie. a total of 675 compartments each averaging about 1,500 ha in size), and the habitat supply models are calibrated to treat each compartment as one unit. As a consequence explicit consideration of the effects of scale on the model outputs is not possible. In the on-going Hinton work, for example, a range of different faunal species have been assessed. These species forage over, and respond to environmental and key habitat factors operating at, spatial scales varying by as much as two orders of magnitude (ie. extending from a few metres to kilometres);

(4) the minimum inputs for HSM are data on the likely forest management prescriptions, forest resource inventory (FRI) data, information on how forest stands change over time and in response to management, and some rules to relate FRI data to habitat requirements for wildlife. FRI data are typically provided by provincial agencies and forest company data-bases and describe commercial timber attributes only. Rules relating FRI data to habitat requirements appear to be developed largely from descriptive studies on the habitat requirements of target species, as reported in the literature; and

(5) the models employed have been largely non-spatial although modellers are developing spatial projection tools for operational use.

## Process-based modelling

A process-based approach to spatial and temporal biophysical modelling has been evolving in Australia over the past two decades (Nix and Gillison 1985, Mackey and Bayes 1990, Norton 1990, Norton *et al.* 1990). This approach is based on spatially-referenced and complementary data sets of primary site-attribute data (eg. climate, terrain, substrate) which are employed within a GIS-framework to estimate the spatial and temporal dynamics of key biophysical processes, such as moisture, energy and nutrient regimes, and to model how they effect biological patterns in space and time (Norton *et al.* 1992).

A variety of levels of spatial resolution are available within the GIS. Gridded data on terrain and climate and hydrology, for example, are available for the Australian continent at a grid resolution of approximately 2.5 km x 2.5 km (Hutchinson and Dowling 1991). Gridded data are available for several smaller but nonetheless relatively large regions (eg. > 1,000,000 ha) at a grid resolution of approximately 200 m x 200 m (eg. Mackey *et al.* 1988), while finer spatial resolution modelling can be undertaken down to a grid size of some 10 m x 10 m, depending on the quality of topographic data (Moore *et al.* 1991, Norton and Moore 1991). The key point is that these techniques could be used in any terrestrial environment and on any continent. Much of the recent research has focussed on those regions in Australia where biophysical data are held at the finest spatial resolution so as to demonstrate the utility of the approach and to refine and compare different models under various tasks. Research in this area is under active development.

The data sets of primary site-attribute data have been used for a variety of biological inventory (see Norton *et al.* 1992), conservation reserve evaluation (Mackey *et al.* 1989) and wildlife modelling tasks. In the case of the latter, several approaches have been adopted depending on the level of knowledge of the habitat requirements of taxa. Where knowledge is extremely limited, the biophysical data sets can be employed to identify sites (ie. grid cells) most likely to support potential habitat. This information can then be used to design more efficient field surveys for the species of interest.

As knowledge of the distribution and habitat requirements of the species improves, the biophysical data sets can be employed with other site data to develop models capable of predicting habitat quality.

These models might take the form, for example, of generalised linear models (GLMS) and logistic regression equations where the presence and abundance of species may be predicted in space and time (Norton and Williams 1989, Norton *et al.* 1990, Lindenmayer *et al.* 1990, Neave and Norton in press).

Illustrated here is the application of such data sets to help identify the potential habitat of the rare and endangered Long-footed Potoroo, which is a nocturnal marsupial restricted to the native eucalypt forests and wet sclerophyll/rainforest ecotones of south east Australia. Knowledge of the distribution and habitat requirements of the species was exceptionally limited at the commencement of the study.

## Identification of Long-footed Potoroo habitat

The long-footed Potoroo is one of the rarest mammals in Australia and is considered to be extremely endangered. Prior to the present study, populations of the species had only been recorded at 34 locations and fewer than 30 specimens had been encountered. Since its discovery in 1978, the species has been at the centre of a controversial forestry debate. The majority of localities supporting populations occur within publicly-owned forests subject to intensive forestry, road construction and fuel-reduction burning.

The immediate task was to identify forest sites most likely to support potoroo habitat so that a comprehensive survey could be undertaken to determine the conservation status of the species and make recommmendations on forest management. A heirarchial approach involving five general steps was adopted to locate potential habitat:

(1) characterise the potential climatic range and, thus, geographic range of the species using the Bioclimatic Prediction System (BIOCLIM; see below for description);

(2) within the potential geographic range of the species predicted by BIOCLIM, exclude those sites which do not continue to support forest cover and thus are unlikely to support potential habitat;

(3) of the forest sites identified in (2), exclude those which are known to have been subjected to significant human-induced disturbances (eg. clearance to exotic pine plantations, clearfelling) and recent high-intensity wildfire;

(4) using the sites not excluded by (1) - (3), above, identify those most likely to support potential habitat on the basis of existing knowledge of the species' habitat requirements. In this case, potential sites were selected for sampling on the basis of their occurrence in gullies and proximity to riparian areas within damp sclerophyll forest, wet sclerophyll forest and rainforest; and

(5) limit sample sites to those forest sites meeting the above requirements and whose combined, contiguous area exceeds 12 hectares (ie. equivalent to three 200 m x 200 m grid cells). This was done to exclude small areas of forest which were considered unlikely to support resident populations.

BIOCLIM is a quantitative method of characterising the potential climatic range of taxa. A description of BIOCLIM was detailed by Nix (1986) and a recent application is given in Williams (1991). BIOCLIM requires as input a file of geocoded (latitude, longitude, elevation) point data for the taxa of concern (target taxa). The point data typically represent those localities where a species, for example, has been observed or collected. From these data, BIOCLIM can define the climatic envelope of the localities with the aid of meteorological data represented as climatic surfaces (Hutchinson *et al.* 1984). The package can then generate a map of a nominated region indicating new sites with similar climate attributes to the sites known to support the taxon and, therefore, where the target taxon has the potential to occur. In this study, BIOCLIM was used to calculate 7 bioclimatic indices for aspects of each of the temperature, precipitation and radiation regimes of the localities where the Long-footed Potoroo had been recorded. These bioclimatic indices summarise the annual and seasonal mean conditions, extreme values and intra-year seasonality in climate experienced at these sites.

Prediction of the potential geographic distribution of the species is based on matching selected thresholds and limits for each of the bioclimatic indices across a grid of points that have climate data. The relatively high spatial-resolution gridded climate data base developed by Richards *et al.* (1990) was used. The grid area is bounded by the co-ordinates 149.00 degrees East and 36.30 degrees South and the south Pacific Ocean. The area is approximately one million hectares in size and spans just to the north of Bega in the State of New South Wales and to the east of Orbost in the State of Victoria. The grid has a spacing of 9 seconds which equates to grid cells of about 200 m x 200 m in size. Thus, for this region, climatic data are available for some 180,000 grid points. The climate at a grid point was considered suitable for the species if it fell within the total percentile range (0-100) for all indices (Nix 1986).

A spatial estimate of extant forest cover for the study region was derived from LANDSAT MSS satellite imagery and was classsified and checked by several procedures including aerial photography and local knowledge (Richards *et al.* 1990). These data were used to exclude sites identified to fall within the potential climatic range of the Long-footed Potoroo but which do not continue to support forest cover due to clearing and so on. The final selection of sites most likely to support potential habitat for the species was based, wherever possible, on knowledge of the spatial distribution of wetter forest types which are commonly associated with potoroo localities (Scotts and Seebeck 1989, Robinson 1991).

A climatic profile of the Long-footed Potoroo was derived using BIOCLIM. The mean (total) annual temperature of sites calculated to be climatically-suitable for the species is 12.4 degrees Celcius, with a mean minimum of 8.9 degrees Celcius and a mean maximum of 13.9 degrees Celcius. The precipitation regime is: mean (total) of 1094 mm with a mean minimum of 916 mm and a mean maximum of 1680 mm. The elevation of sites known to support populations of the species ranges from 100 m to 1080 m above sea level. The average elevation is 427 a.s.l.

The predicted potential distribution of the Long-footed Potoroo within extant forest cover in south east New South Wales was plotted. Less than 20% of sites predicted to be climatically-suitable by BIOCLIM have been cleared of their forest cover by humans. Overall, some 3,300 grid cells within forest are predicted to be climatically-suitable. This equates to a total area of approximately 13,200 hectares, which falls into three land tenure categories: about 70% on public land which is subjected to intensive logging, about 25% in reserves, and about 5% on private land where relatively few management prescriptions currently apply. On the basis of the additional site-selection criteria (eg. forest association, disturbance), some 1,700 (6,830 ha) of the 3,300 grid cells were considered most likely to support potential habitat. Since 1988 about 145 of the sites have been harvested, representing a modification of some 450 hectares of potential habitat. Another 255 sites (1,020 ha) were scheduled for logging over the next year.

On the basis of these analyses, forestry precriptions and harvesting schedules were modified so that the sites predicted most likely to support potential habitat for the potoroo could be intensively surveyed before harvesting.

Field sampling is currently in progress and already six sites identified using this approach have been confirmed to support the species.

## Discussion

The case study on predicting the location of potential habitat of the Long-footed Potoroo helps to illustrate the utility of a process-based approach to wildlife modelling. Despite the extremely limited understanding of the ecology of the species, the ability to estimate and manipulate site-attribute data at a relatively fine spatial resolution enabled the rapid identification of sites most warranting survey within a relatively large forest region (ie. about 0.15% of the total area). As more data become available on the distribution and abundance of the animal in relation to other site attributes (eg. slope, aspect, index of soil wetness) contained in the GIS data-base, it will be possible to develop a model of the processes most strongly influencing habitat quality and to predict spatial variation in habitat quality at a fine scale. With knowledge of the likely effects of different management scenarios on these processes, it will also be possible to predict the dynamics of habitat quality through time.

In contrast, current approaches to habitat supply analysis and modelling generally use pre-classified forest inventory data as a template for modelling wildlife habitat and wildlife population dynamics. Some of the more important assumptions inherent in these approaches are (1) that the type of inventory data collected are relevant to, and sufficient for, evaluating the habitat requirements of wildlife species; (2) that the timing of inventory and the scales that inventory data were collected/derived at are relevant to, and sufficient for, evaluating the habitat requirements of wildlife species; and (3) that the design and conduct of the forest inventory used to generate a classification was systematic and the location of sample sites was sufficiently representative to ensure an adequate coverage of the biophysical variation of the target region. These assumptions are unlikely to be valid for most wildlife and at least one province is contemplating the development of ecologically (biophysical) based forest inventories to support habitat supply modelling.

Clearly, significant opportunities exist to adapt and unify the best concepts underpinning both the HSA/HSM and biophysical approaches to wildlife modelling, thereby improving the reliability of model predictions and the quality of integrated forest management in North America and Australia. This would certainly be the case for temporal predictions concerning changes in habitat quality or the wildlife carrying capacity of various forest types under different harvesting prescriptions and scenarios. The biophysical modelling approach could, for example, be used to characterise the key processes effecting changes in habitat quality and to determine the most appropriate scales of aggregation to undertake habitat and population simulations.

Since model simulations through time are largely untestable in the short term, it will be essential to undertake monitoring of key species, key habitats and key processes at strategic places and times in the forests. A process-based modelling approach using the types of site-attribute data described above can help to select these parameters and could be expected to make the monitoring process more cost-effective and informative.

## Acknowledgements

This collaboration, and part of the research presented, was initiated while TWN was a visiting scientist at Forestry Canada (Ontario Region), Sault Ste Marie during June-July 1991. We thank Forestry Canada and the Ontario Ministry of Natural Resources for their support. The research was supported also by the Australian Research Council and the New South Wales National Parks and Wildlife Service.

## References

Allen, A.W. (1982). **Habitat Suitability Index Models: Marten.** FWS/OBS-82/10.00, Biological Services Program, Fish and Wildlife Service, U.S. Department of the Interior, Washington, D.C. 8 pp.

Allen, A.W., Jordan, P.A. and J.W. Terrell (1987). **Habitat Suitability Models: Moose, Lake Superior Region.** Biological Report 82(10.155), Fish and Wildlife Service, U.S. Department of the Interior, Washington, D.C. 47 pp.

Bonar, R., Quinlan, T., Sikora, T., Walker, D. and J. Beck (1990). **Integrated timber and wildlife management on Weldwood's Hinton Forest Management Area.** Unpublished manuscript, Weldwood of Canada Ltd. Hinton, Alberta.

Boyce, S.G. (1980). **Management of forests for optimal benefits (DYNAST-OB).** USDA Forest Service Research Paper SE-204. South eastern Forest Experimental Station. 92 pp.

Covington, W.W., Wood, D.B, Young, D.L., Dykstra, D.P. and L.D. Garret (1988). TEAMS: a decision support system for multiresource planning. **Journal of Forestry** 86: 25-33.

DeGraaf, R.M. and D.D. Rudis (1986). **New England wildlife: habitat, natural history, and distribution.** USDA For. Serv. Gen. Tech. Rep. NE-108. Northeastern For. Exp. Sta. 491 pp.

Duinker, P.N. (1986). **A Systematic Approach for Forecasting in Environmental Impact Assessment: A Deer Habitat Case Study.** Unpublished Ph.D. Thesis, Faculty of Forestry, University of New Brunswick, Fredericton, New Brunswick.

Duinker, P.N. (1989). **Wildlife in forest management: from constraint to objective.** pp. 133-144 in Symposium Proceedings, Forest Investment: A Critical Look. Great Lakes Forestry Centre, Forestry Canada, Sault Ste Marie, Ontario.

Duinker, P.N., Higgelke, P. and S. Koppikar (1991). **GIS-based habitat supply modelling in Northwestern Ontario: moose and marten.** Paper presented at GIS '91, Vancouver, British Columbia.

Greig, L.A. Duinker, P.N., Wedeles, C.H.R. and P. Higgelke (1991). **Habitat Supply Analysis and Modelling: State of the Art and Feasibility of Implementation in Ontario.** Prepared for Ontario Ministry of Natural Resources. ESSA Environmental and Social Systems Analysts Ltd. Report No. 91-403.2. 81 pp.

Hutchinson, M.F, and T. Dowling (1991). A continental hydrological assessment of a new grid-based digital elevation model of Australia. **Hydrol. Proc.** 5: 45-58.

Hutchinson, M.F., Booth, T.H. McMahon, J.P. and H.A. Nix (1984). Estimating monthly mean values of daily total solar radiation for Australia. **Solar Energy** 32: 277-290.

Lindenmayer, D.B., Cunningham, R.B., Tanton, M.T. Smith, A.P. and H.A. Nix (1990). The conservation of arboreal marsupials in the montane ash forests of the central highlands of Victoria, south east Australia. I. Factors influencing the occupancy of trees with hollows. **Biol. Conserv.** 54: 111-131.

Mackey, B.G., Nix, H.A., Hutchinson, M.F. McMahon, J.P. and P.M. Fleming (1988). Assessing representativeness of places for conservation reservation and heritage listing. **Environ. Manage.** 12: 501-514.

Mackey, B.G., Nix, H.A. Stein, J.A., Cork, S.E. and F.T. Bullen (1989). Assessing the representativeness of the Wet Tropics of Queensland World Heritage Property. **Biol. Conserv.** 50: 270-303.

Mackey, B.G. and A. Bayes (1989). A modelling framework for the spatial extension of ecological relations in vegetation studies. pp. 93-97 in **Proceedings of the Eighth Biennial Conference of the Simulation Society of Australia** IMACS, Canberra.

Marcot, B., McNay, R.S. and R.E. Page (1988). **Use of Microcomputers for Planning and Managing Silviculture Habitat Relationships.** General Technical Report PNW-GTR-228, Pacific Northwest Forest Experiment Station, USDA Forest Service, Oregon. 19 pp.

Martelle, M. (1990). **Potential Responses of Marten Habitat Carrying Capacity Effects to Future Timber Harvests in a Northern Ontario Forest.** Miscellaneous Report No. 06, Northwestern Ontario Forest Technology Development Unit, Ministery of Natural Resources, Thunder Bay, Ontario. 80 pp.

Moore, I.D., Grayson, R.B. and A.R. Ladson (1991). Digital terrain modelling: a review of hydrological, geomorphological and biological applications. **Hydrol. Proc.** 5: 3-30.

Moore, T.G.E and G.C. Lockwood (1990). **The HGS wood supply model: description and user's manual.** Forestry Canada Inf. Rep. PI-X-98. Petawawa National Forestry Research Institute, Chalk River, Ontario. 31 pp.

Neave, H.M. and T.W. Norton (in press). Integrated management of forest wildlife: new ways to research habitat. in **The Conservation and Management of Australian Forest Fauna** (ed. D. Lunney). Royal Zoological Society of New South Wales and Surrey Beatty and Sons, Sydney.

Nix, H.A. (1986). A biogeographic analysis of Australian elapid snakes. pp. 4-15 in **Atlas of Elapid Snakes of Australia** (ed. R. longmore), Australian Flora and Fauna Series No.7 AGPS, Canberra, Australia.

Nix, H.A. and A.N. Gillison (1985). Towards an operational framework for habitat and wildlife management. pp. 39-55 in **Wildlife Management in the Forests and Forestry-Controlled Lands in the Tropics and Southern Hemisphere** (ed. J. Kikkawa). University of Queensland Printery , Brisbane, Australia.

Norton, T.W. (1990). Maintaining biological diversity. pp. 53-74. in **Climate change in Australia: environmental, socioeconomic and political considerations**. (eds. T.W. Norton and O.B. Williams). CRES, The Australian National University, Canberra, Australia.

Norton, T.W. and I.D. Moore (1991). Ecological applications of GIS and hydrological modelling: identification and management of Koala habitat in south east Australia. pp. 637-648 in **Proceedings of International Hydrological and Engineering Conference**, Perth, November.

Norton, T.W. and J.E. Williams (1990). Wildlife management in Australia: new developments and opportunities using computer-based generic models. **Math. Comp. Simul.** 32: 231-235.

Norton, T.W., Mackey, B.G. and D.B. Lindenmayer (1990). Comments on biological and environmental data sets required for the Australian National Forest Inventory. **Aust. For.** 53: 88-95.

Norton, T.W., Neave, H.M., Nix, H.A. and I.D. Moore (1992). Approaches to forest survey, inventory, evaluation and planning using biophysical models of landscape processes. **Conference Proceedings, GIS '92**, Vancouver, British Columbia, Canada.

Patch, J.R. (1987). Habitat supply analysis and forest management. **Transactions, North America Wildlife and Natural Resources Conference** 52: 53-59.

Richards, B.N., Bridges, R.G., Curtin, R.A., Nix, H.A., Shepherd, K.R. and J. Turner (1990). **Biological Conservation of the South-East Forests**. AGPS, Canberra, Australia.

Robinson, R.W. (1991). **An analysis of the vegetation at ten sites where the Long-footed Potoroo has been recorded**. Unpublished report to the Victorian Department of Conservation and Environment, Melbourne, Victoria.

Scotts, D. and J.H. Seebeck (1989). **Ecology of Potorous longipes (Marsupialia: Potoroidae); and preliminary recommendations for management of its habitat in Victoria**. Arthur Rylah Institute for Environmental Research, Technical Report Series No. 62, Victorian Department of Conservation, Forests and Lands, Melbourne, Victoria.

Thomas, J.W. (1979). **Wildlife Habitats in Managed Forests: the Blue Mountains of Oregon and Washington**. Agricultural Handbook No. 553, USDA Forest Service, Washington D.C. 512 pp.

Verner, J., Morrison, M.L. and C.J. Ralph (1986). **Wildlife 2000: Modelling habitat relationships of terrestrial vertebrates**. University of Wisconsin Press. 470 pp.

Wang, E., Erdie, T. and T. Roussell (1987). **FORMAN wood supply model: user manual**. New Brunswick Executive Forest Research Committee Inc. Fredericton, New Brunswick. 61 pp.

Watt, W.R. (1990). **A Program of Technology Development for Wildlife Habitat Management in Northern Region Forests**. Northern Forest Development Group, Ministry of Natural Resources, Timmins, Ontario. 30 pp.

Williams, J.E. (1991). Biogeographic patterns of three sub-alpine eucalypts in south-east Australia with special reference to *Eucalyptus pauciflora*, Sieb. ex Spreng. **J. Biogeog.** 18: 223-230.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# B.C. Indian fisheries information system: communication techniques for consultation - negotiation interim report

**Diane Evans**

Contractor
Federal Department of Indian Affairs
7504 - 148 Street,
Surrey British Columbia, V3S 3E7;
Tel: (604) 596-8931

**Joe Kuhn**

Comprehensive Claims Branch,
Department of Indian Affairs and Northern Development,
Les Terrasses de la Chaudiere,
Hull, Quebec, K1A 0H4;
Tel: (819) 953-2771.

## Abstract

*The work of Canada's Department of Indian Affairs and Northern Development (DIAND) in aboriginal fisheries policy and community support increasingly falls in the area of communications and consultation. This is especially true in B.C. where First Nations, federal and/or provincial governments will be involved in watershed/sea-zone Indian fishery management areas (IFMAs) and the negotiation of approximately 30 comprehensive claims agreements (treaties), each with fisheries elements.*

*As federal manager of fisheries, the Department of Fisheries and Oceans (DFO) has the lead role in developing fisheries agreements with First Nations. DIAND supports both DFO and First Nations in this process. When it comes to treaty negotiations, however, it is DIAND which represents the federal government. Thus, DIAND has an interest in the methods of fisheries information exchange between parties (including the B.C. government), and how they can be used to support consultation, negotiation and joint decision-making.*

*This presentation reports on the development of an initial design for a model non-technical fisheries information system. It focusses on three areas of B.C. and utilizes a software program called Quikmap. Base maps for these areas are being obtained or created and sources of thematic information (data and interpretive information) are being identified and utilized. Such an information system will be useful to DIAND staff, IFMA workers and, especially, treaty negotiation teams. Private sector interest groups and teachers/students will eventually have opportunity to use the system and participate in its further development, especially at the local level.*

## Introduction

The project, "B.C. Indian Fisheries Information System: Communication Techniques for Consultation-Negotiation", was contracted to Diane Evans in Vancouver, reporting to Joe Kuhn, Comprehensive Claims Branch, Department of Indian Affairs and Northern Develoment (DIAND), Ottawa. Its duration is approximately six months and will be completed by the end of February, 1992. This is an interim report, therefore, briefly describing work-in-progress. No final conclusions have yet been drawn.

The work of DIAND in aboriginal fisheries policy and community support increasingly falls in the area of communications and consultation. This is especially true in B.C. where First Nations, federal and/or provincial governments will be involved in watershed/sea-zone Indian fishery management areas (IFMAs) and the negotiation of approximately 30 comprehensive claims agreements (treaties), each with fisheries elements.

As federal managers of fisheries, the Department of Fisheries and Oceans (DFO) has the lead role in developing fisheries agreements with First Nations. DIAND supports both DFO and First Nations in this process. When it comes to treaty negotiations, however, it is DIAND which represents the federal government. Thus, DIAND has an interest in the methods of information exchange between parties (including the B.C. government), and how they can be used to support both negotiation and joint decision-making.

To this end it is necessary to consult with information managers and negotiators in the three orders of government, as well as knowledgeable non-government people, to determine techniques currently used for managing resource information, assess how well this information is shared amongst participants, and to gain some understanding of user needs. What we are looking for is a low-cost, user-friendly means by which negotiators in the three orders of government might successfully share natural resource information with each other at the bargaining table.

## Objective and scope

The main objective of the project is to investigate three or more watershed/sea-zone fishery areas in B.C. as a basis for designing a preliminary demonstration model for an information system especially suited to the fisheries negotiation process. The term "information system", as used in the context of this project, is broader than the conventional GIS use of the term. Besides using a computer desktop mapping program called Quikmap, other graphic communication techniques are also considered. These are not discussed in this paper, however.

The project will obtain or create base maps for the study areas, identify relevant sources of information, and select portions of data and interpretive information which best serve to demonstrate (to non-technical users, especially) how the system can input, store, retrieve and graphically display information. It is anticipated that information on the following aspects of IFMAs will be required for fisheries negotiations, but the study may modify these points:

1. key biophysical information *i.e.* the "what" and "where" of water bodies and fishery resources;
2. information such as reserves, claim boundaries, traditional Indian fishing sites, industrial sites, etc;
3. other fisheries in the IFMA (*e.g.* commercial, sport), noting conflict relationships;
4. environmental impact considerations of major programs or projects in the watershed/sea-zone areas under study, and resulting conflicts;
5. present fisheries and habitat management structure and process: how and by whom are management decisions made;
6. proposed fisheries and habitat management structure and process: alternatives/advocates;
7. current Indian community human resource development programs receiving federal or provincial government support and the specifics of that support relative to fisheries development.

## What is Quikmap?

Quikmap, developed by AXYS Software of Sidney, B.C., is a desktop mapping system for IBM PC/AT compatible computers. It is referred to as a "low-end" GIS because of its relatively low analytical capability and cost. It offers significant advantages over conventional "high-end" GIS because (1) it is easy to use (proficiency can be attained in a matter of weeks/months as compared to months/years for many "high-end" GIS); (2) data are not fundamentally linked with the base maps, which allows data to be displayed on different maps with different scales; (3) graphic displays/data files may be exchanged between different users with ease; (4) custom menus can be created, allowing computer novices access to map information; (5) price/performance is unequaled in the GIS market with costs around $1000 per workstation; and (6) Quikmap runs on any IBM-PC compatible with graphics and a hard drive.

Quikmap workstations can be used to access information prepared using more complex GIS systems. Detailed data analysis can be done on a more powerful system such as Arc/Info or Pamap, then summarized and downloaded to Quikmap in a PC environment. Optional translators will import Autocad (.DXF), Intergraph (.DGN) and Caris (.NTX) files.

## Work-in-progress

The first phase of the current study was to identify federal, provincial and First Nations projects with similar objectives and assess problems and promising results. From this, three intensive effort areas were selected: the Cowichan, Nimpkish and Deadman River watersheds.

An investigation was conducted to locate any digital base maps and geo-referenced thematic information already in existence for the above areas, but it soon became apparent that there is a general lack of inexpensive, readily available information of this type for given areas of the province. To compound the problem there also seems to be a lack of agreed-upon map and data standards amongst government agencies, and a range of incompatible digital formats amongst users of GIS.

For these reasons, as well as limited time, budget and computer hardware (*e.g.* digitizing tablet), aspects of several GIS and/or desktop mapping projects already underway are being considered for the purposes of this project. Arrangements have been made with the various federal/provincial government agencies and First Nations groups involved in the three study areas to utilize samples of their data and maps ("windows") in order to produce a demonstration model to meet the objectives of this project.

Within each study area there is some degree of cooperative sharing of information with the following groups and their respective projects:

1. Cowichan River watershed, near Duncan, Vancouver Island
   - the Fisheries Information System (FINS) is a prototype fisheries production management software program that uses the Cowichan River watershed as a study area. It is a database management shell built around Quikmap. Development of the system, by AXYS Software of Sidney, B.C., was sponsored by the Department of Fisheries and Oceans (DFO) and B.C. Ministry of Environment, Lands and Parks (MOELP) as part of their joint Fish Habitat Inventory and Information Program.
   - the Cowichan Band is involved in developing further, with LGL Ltd, Sidney, B.C., a FINS-type system for their own use, using Arc/Info then downloading to Quikmap.
   - B.C. Ministry of Environment, Lands and Parks (MOELP), Environmental Emergency Services Branch, Victoria, and contractor LGL Ltd., Sidney, B.C.
   - MOELP recently awarded LGL Ltd contracts for the compilation and sensitivity mapping of biological resources, human use and special status areas for the Southern Strait of Georgia and design and production of an oil spill response atlas. It is hoped that "map samples" of the Cowichan area can be downloaded to Quikmap and used in this project.
2. Nimpkish River watershed, near Port Hardy, Vancouver Island
   - Musgamagw Tribal Council (MTC) of Alert Bay is involved in a project which utilizes a newly-developed land-based resource management shell, which AXYS Software has built around Quikmap. The shell encompasses all watersheds in the MTC area from Nimpkish on Vancouver Island to Kingcome on the mainland.
3. Deadman River watershed, draining into the Thompson River west of Kamloops
   - the Deadman River is in the traditional territory of the Shuswap Nation Tribal Council (SNTC). It was selected because the Skeetchetsn Band and SNTC are involved in a community-based watershed management committee, as well as a co-management initiative with DFO. (The Bonaparte River watershed, west of the Deadman, is also of some interest due to the Bonaparte Band's participation in an SNTC pilot project using Quikmap.)

In addition to the study areas, information from B.C. MOELP, Recreational Fisheries Branch, Victoria, is used in this project. They also have a Quikmap demonstration project which "windows" various areas in each Region of the province to demonstrate the capabilities of Quikmap in terms of regional fisheries information needs. (For more information on this topic, see Peter Lewis's paper in these proceedings).

Many First Nations organizations already have or are in the process of acquiring GIS of their own, whether or not they are actively involved in comprehensive claims, specific claims, fisheries co-management agreements, or other negotiation considerations. Leaders at both the band and tribal council level see GIS as a necessary tool in effectively managing the land, water and other resources in their traditional territories. Also, it is assumed by some that GIS enables First Nations to consult with various government agencies "on their own terms".

## Interim conclusions

For federal, provincial and First Nations governments to successfully reach formal fisheries agreements and treaties, work group committees and negotiating teams must have an orderly method for gathering, storing, and displaying key resource information that is relevant to the process. Negotiators have limited use for complex statistics, large databases, myriads of technical reports, machine modelling and analyses. What they require is a synthesis of this information, a general overview of the resources in an area such as a watershed, and their use and importance in the past, present and future.

The power of graphic images cannot be overestimated in the above process because, as we all know, one gets so much more information in the blink of an eye than one can verbalize in hours. For instance, thematic information overlain on a base map and combined with other thematic information can visually illustrate various relationships and values via shapes, symbols, sizes, colours, patterns, etc. Such illustration, using Quikmap, may be sufficient to focus discussion on easily recognized and generally accepted ecosystem elements, taking into account integrated resource management principles and First Nations traditions.

Further computer analysis using a "high-end" GIS may not be necessary because key/summary information, in primarily graphic form, empowers users to better utilize their individual intuitive and assessment skills to do their own mental analyses during consultations. An inexpensive, easily learned desktop mapping program such as Quikmap is thus part of a structure designed to generate and process an information flow to aid decision-making around the bargaining table.

This project is mainly an investigative mission and learning experience for participants, not only in terms of GIS and "state-of-the-art" communications technology, but also in terms of First Nations aspirations and the changing roles of DIAND and DFO. Many thanks are extended to those kind and enthusiastic individuals who have helped with this project, for ultimately it is their project, too.

*Any comments, suggestions or questions will be appreciated.*

# C2 GIS for habitat modelling

**Continuous habitat suitability modelling using convolution filters**

John Rickers, University of Minnesota, St. Paul, MN, USA

**The Ecosystem Management Model project: integrating simulation modelling and GIS in the Canadian Parks Service**

David Buckley et al., TelTec GIS Division, Longmont, CO, USA

**Using GIS to analyze Spotted Owl habitat within Oregon**

Pat Green, Bureau of Land Management, Portland, OR, USA

**Managing and monitoring wildlife habitat in the North Yukon *poster***

Manon Desforges et al., Fish and Wildlife Branch, Whitehorse, YT, Canada

G I S
92

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Continuous habitat suitability modeling using convolution filters

**John Rickers**

Department of Forest Resources
University of Minnesota
115 Green Hall
N. Cleveland Avenue
St. Paul, MN 55108

## Abstract

*Wildlife habitat models, in particular Habitat Suitability Indices (HSI) have been limited in their application. Traditionally, these models tend to be very data intensive and site specific. Thus, most HSI models cannot be supported by traditional forest inventory, and are therefore not amenable to application over large areas. Also, until the inception of Geographic Information Systems (GIS) there was a lack of a tool which would efficiently consider the implicit spatial parameters of such models. The HSI modeling approach discussed in this paper attempts to tackle these traditional problems. A raster-based GIS was used to implement a compartment-based HSI model for ruffed grouse (Bonasa umbellus). It has been designed for areas where quaking aspen (Populus tremuloides) is both abundant and commercially important and is the essential habitat type providing the life requirements of ruffed grouse. Since this model uses traditional forest inventory data, it may be applied over large areas where aspen holds these values. This approach models the spatial distribution of different age classes of aspen and how this distribution fulfills the year-round life requirements of the ruffed grouse. It is continuously applied through the use of convolution filters (moving windows).*

## Introduction

In the past wildlife habitat models have been limited in their application. Typically, models have been developed to increase understanding of the relationships between a particular species and its habitat. After, they are either not applied, or are applied on a very site-specific basis. This lack of application generally stems from three reasons: 1) a limited understanding of species/habitat relationships, 2) the absence of data to support individual models, and 3) the need for a tool to efficiently analyze implicit spatial parameters.

With the inception of Geographic Information Systems (GIS), new developments in wildlife habitat modeling have taken place. In particular, the application of Habitat Suitability Indices (HSI)(USFWS 1981) have made great advancements (Pereira 1990; Mead et al. 1988; and Donovan et al. 1987). The fact that HSIs are based on assessing the capability of providing a particular species life requirements in a specified area, lends itself in particular to implementation in a raster-based system.

This paper discusses a model which has evolved from extensive research on ruffed grouse and aspen where species/habitat relationships are well developed. It is based on traditional, stand-mapped forest inventories and can be applied where aspen is the critical resource for ruffed grouse. It has been implemented in a raster-based GIS which affords the ability to efficiently analyze inherent and important spatial parameters.

## Ruffed grouse and aspen

Ruffed grouse are an important biological as well as economical resource in Minnesota. It is a token of Minnesota's forested lands. It is the State's most abundant game bird, and is claimed by many hunters as the premier upland game bird. Revenues of approximately 10.5 million dollars (US) are procured by the State annually from the hunting of this species: this figure approaches 30 million dollars during peak years (MNDNR 1987).

The importance of aspen in providing the highest quality food and cover for ruffed grouse in Minnesota has been well documented (Gullion 1977; Gullion 1984; Gullion and Svoboda 1972; Svoboda and Gullion 1972; Back 1982; and Buech et al. 1990), and its importance has also been noted in Wisconsin (Kubisiak 1978; Kubisiak 1985; and Kubisiak et al. 1980). Aspen is also a very important economical resource in Minnesota. Approximately 2.4 million cords of aspen are removed from Minnesota forests annually and this figure continues to rise (Jaakko Poyry Consulting 1991).

Because aspen, and in particular an interspersion of equal distribution of different ages, is important to ruffed grouse and because there is a substantial demand for this timber type, it is clear that timber management activities can affect grouse habitat. By dispersing aspen harvests both spatially and temporally, thus providing interspersion of equal age distributions over an area, habitat for ruffed grouse can be enhanced while at the same time achieving harvesting goals.

## Ruffed grouse habitat suitability index

This model is based on the assumption that ruffed grouse have four basic life requirements that must be found within a 10 acre evaluation area (male ruffed grouse activity center). These life requirements can be found in four different age classes of aspen.

The following four age classes were taken from Buech et al. 1990:

| **Life requirement** | **aspen age class** |
|---|---|
| Brood Cover | 0-12 Years Old |
| Fall-Spring Cover | 13-25 Years Old |
| Nesting Cover and Food | 26-38 Years Old |
| Winter Food | 39+ Years |

The model developed was also based on the following assumptions:

1) optimal habitat will be found where the evaluation area consists of 25% of each of the four age classes.

2) where ruffed grouse have suitable food and cover, they have the potential to increase and develop stability in the absence of other mortality factors.

The following formulation is used to assess the habitat quality of an evaluation area:

$$HSI = (SIV1 + SIV2 + SIV3 + SIV4)/4$$

where;

SIV1 = (f) percent of area in 0-12 year aspen

SIV2 = (f) percent of area in 13-25 year aspen

SIV3 = (f) percent of area in 26-38 year aspen

SIV4 = (f) percent of area 39+ year aspen

*Figure 1 is the curve function used for determining each SIV

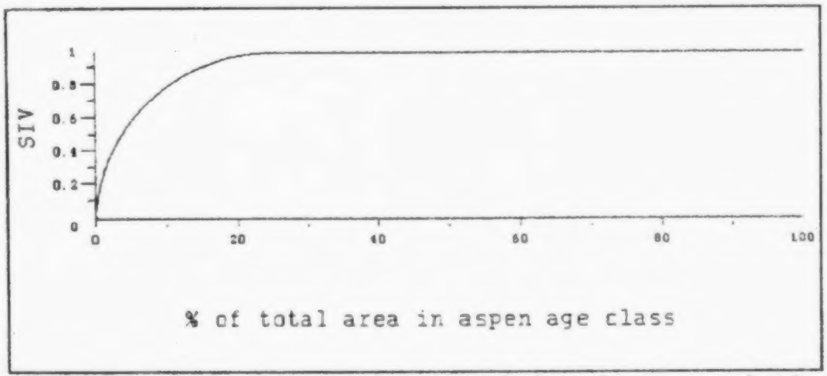


**Figure 1.** Derivation of suitability index variable for each aspen age class

## Study area

The study area is a two-township area in north-central Minnesota which was chosen for use in a multi-resource integration project at the University of Minnesota (see Queen 1992) . Mapped stand-level inventories were obtained from all ownerships within the area and digitized in ARC 3.4D. Stand-level databases were also obtained and integrated to create a comprehensive, ownership-independent database for the area (see Arthaud 1992).

## GIS modeling approach

The ARC 3.4D coverage was rasterized using a 1 acre minimum mapping unit and imported to ERDAS 7.4 for model implementation. A descriptor file was set up with cover type and age for each of the stands. A recoding procedure was used to determine the age class which each aspen stand fell into.

The implementation approach discussed below involves the use of convolution filtering (also referred to as "moving window"). This technique examines the class values of cells of a specified square area or window and assigns a value to the center cell of that window based on a user-indicated option such as a total of class values, the highest class value, or an average of class values.

Since this technique is limited to square convolution filters, meaning odd heights and widths, a 5x5 (25 acre) filter was chosen for this implementation. Alternatively, a 3x3 (9 acre) filter could have been used but; this was below the specified minimum habitat area (10 acres). This filter was then moved across every row in the study area and counts of the cells of each of the age classes in the filter was assigned to the center cell of the filter, thus achieving counts for each cell in the area. The modeling algorithm (from the formulation above) was written using ERDAS' GIS modeling language and HSI values were calculated and then assigned to each cell in the area.

## Results

The HSI model which was developed for this project is quite simple and can be used where mapped, stand-level forest inventories are available, thus making it very usable. Automation of models such as these can be automated in a GIS making analysis much more efficient. Output maps are no longer stand-based; rather mapped units become habitat-based and consist of complex features composed of combinations of portions of stands from different aspen age classes in or at different proximities. In essence this modeling approach reflects the distribution of aspen in an evaluation area with respect to size of aspen stands, their age class, and their interspersion.

## References

Arthaud, G.J. 1992. *Multi-ownership integration of forest stand data: A case study in Itasca County, Minnesota. In* Proceedings GIS '92 Symposium. Vancouver, Canada.

Back, G.N. 1982. Impacts of management for ruffed grouse and pulpwood on nongame birds. Ph.D. thesis, University of Minnesota. 96 pp.

Buech, R.R., C.J. Martin, and H.M. Rauscher. 1990. *The grouse management advisory system compartment module.* USDA. For Serv. North Cent. For. Serv. Exp. Stn. St. Paul, MN.

Donovan, M.L., D.L. Rabe, and C.E. Olson, Jr. 1987. *Use of geographic information systems to develop habitat suitability models.* Wildl. Soc. Bull. 15:574-579.

Gullion, G.W. 1977. *Forest manipulation for ruffed grouse.* Trans. N. Amer. Wildl. Nat. Res. Conf. 42:449-458.

Gullion, G.W. 1984. *Managing northern forests for wildlife.* Minn. Agric. Exp. Stn., Misc. J. Ser. Publ. No. 13, 442. 72 pp.

Gullion, G.W. and F.J. Svoboda. 1972. *The basic habitat resource for ruffed grouse.* Proc. Aspen Symp., U.S.D.A. For. Serv. Gen. Tech. Rep. NC-1:113-119.

Jaakko Poyry Consulting. 1991. *Initial harvesting scenarios for a generic environmental impact statement: Timber harvesting and management in Minnesota. Jaakko Poyry Consulting,* Tarrytown, NY. 45 pp.

Kubisiak, J.F. 1978. *Brood characteristics and summer habits of ruffed grouse in central Wisconsin.* Wisc. Dept. Nat. Res. Tech. Bull. 108. 12 pp.

Kubisiak, J.F. 1985. *Ruffed grouse habitat relationships in aspen and oak forests of central Wisconsin.* Wisc. Dept. Nat. Res. Tech Bull. 151.

Kubisiak, J.F., J.C. Moulton and K.R. McCaffery. 1980. *Ruffed grouse density and habitat relationships in Wisconsin.* Wisc. Dept. Nat. Res. Tech. Bull. 118. 15 pp.

Mead, R.A., L.S. Cockerham, and C.M. Robinson. 1988. *Mapping gopher tortoise habitat on the Ocala National Forest using GIS.* pp. 395-400. *In* Proceedings of GIS/LIS '88. Third Annual International Conference, Exhibits and Workshops. San Antonio, Texas. Nov. 30-Dec. 2, 1988.

Minnesota Department of Natural Resources. 1990. *Long range planning document* - Chpt. 31. Ruffed Grouse. Vol. 2, pages 31-1 — 31-15. Public review draft.

Pereira, J.M.C. 1990. *Habitat suitability analysis and impact assessment with GIS-based logistic multiple regression modeling.* pp. 119-125. *In* Proceedings of GIS '91. Feb. 12-5, 1991. Vancouver, Canada.

Queen, L.P. 1992. *Integrating spatial models of recreation, wildlife, and timber resources. In* Proceedings GIS '92 Symposium. Vancouver, Canada.

Svoboda, F.J. and G.W. Gullion. 1972. *Preferential use of aspen by ruffed grouse in northern Minnesota.* J. Wildl. Manage. 36:1166-1180.

U.S. Fish and Wildlife Service. 1981. *Standards for the development of suitability index models.* Ecological Services Manual 103. U.S. Department of Interior, Fish and Wildlife Service, Division of Ecological Services. Government Printing Office, Washington, D.C.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# The Ecosystem Management Model project: integrating simulation modelling and GIS in the Canadian Parks Service

**David J. Buckley**

TelTec, GIS Division
Suite 210 - 1430 Florida Avenue
Longmont, CO, USA 80501

**Dennis O'Leary**

Jerry Bentz
Geowest Environmental Consultants Ltd.
3614 Hillview Crescent
Edmonton, AB, Canada T6L 1C4

**Chuck Blyth**

Canadian Parks Service
Elk Island National Park
R.R.1. Site 4
Fort Saskatchewan, AB, Canada T81 2N7

## Abstract

*GIS technology has traditionally been used as a descriptive tool in managing land resources. However, with the increasing sophistication of users and software offerings, GIS has the potential of providing predictive capabilities in an operational planning environment. The Canadian Parks service has initiated an Ecosystem Management Model (EMM) strategy to systematically implement GIS technology in this manner. The EMM project will enable Elk Island National Park wardens the ability to explore the consequences of resource management actions prior to their implementation. This project is intended as a prototype for all National Parks in Canada.*

*Elk Island National Park (EINP), situated in the Aspen Parkland of central Alberta, is a completely enclosed game park known for its abundant bison, elk, moose and deer populations. EINP has a long and complex history of environmental management that has resulted in animal populations being managed at near ecological carrying capacity. In 1990, EINP staff began an investigation into the application of simulation modelling for enhancing resource management planning. It was concluded that sufficient data, understanding, expertise and technology exists to construct models that could be used to simulate ecosystem processes and evaluate management decisions. The use of simulation models has been identified as a key goal for the management of both the vegetation and ungulate populations. Simulation modelling is dynamic in nature and supports the pre-dictive assessment of management assumptions, actions and inactions. GIS technology provides the foundation for the implementation of these simulation models.*

*This paper reviews the design and development of the EMM with an emphasis on the issues related to GIS implementation, application development and integration of simulation modelling. Several different models will be presented as examples. As well, the evolution of GIS as a predictive tool, as opposed to a descriptive tool, will be discussed in the context of temporal GIS requirements. In addition, strategies and issues related to customizing the operational GIS environment will be discussed.*

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Using GIS to analyze Spotted Owl habitat within Oregon

**Patrick D. Green**

Bureau of Land Management
Oregon State Office
Portland, Oregon 97208

## Abstract

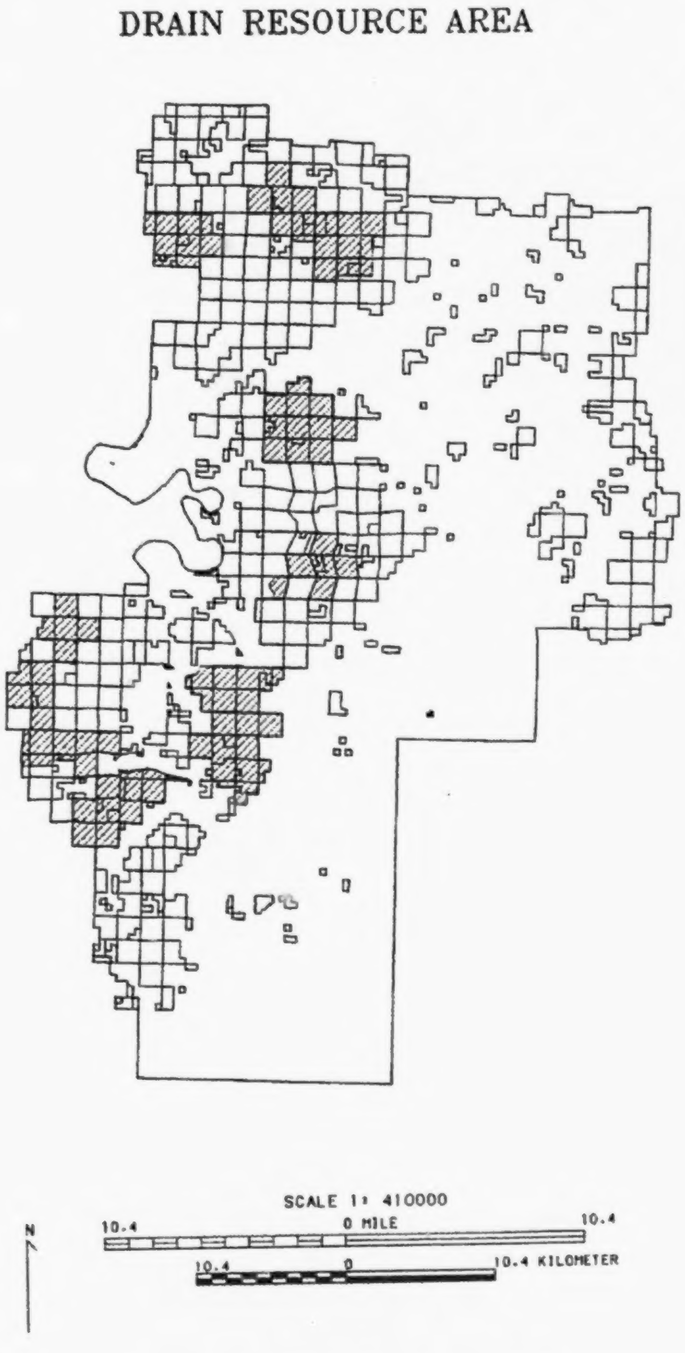
*The United States Department of Interior (USDI), Bureau of Land Management (BLM) built the Western Oregon Digital Database (WODDB) from 1985 to 1990. In April 1990 the Interagency Scientific Committee (ISC) Report, A Conservation Strategy for the Northern Spotted Owl, was released. Analysis of the impact of these conservation areas was the maiden voyage of WODDB, BLM's first attempt to use the new database and GIS technology to address a major spatial issue. Since the release of the ISC report, WODDB has addressed seven different spatial solutions to spotted owl habitat conservation. The database was also used to develop and defend the Bureau's Timber Sale plans given these new owl habitat protection areas. These owl habitat-related projects will be the topic of this paper.*

## Introduction

In the mid 1980s BLM decided to use GIS for the 1990 decadal Resource Management Plans (RMPs) for western Oregon. This decision was based on the increasing spatial complexity of resource values and issues. The BLM Oregon management team realized that the balance between historic use of forests for timber and other resource values couldn't be found by manual methods (grease pencils and dot grids). Therefore, in around 1985 the WODDB project was born. BLM spent the next five years building the database for 2.5 million acres scattered over 500 townships (7 million total acres). The building of the database has been described by Dippon *et. al.* 1989. Today WODDB includes over 100 themes, each covering from a few hundred acres to over 20 million in spatial extent.

During the initial design (1985/86) two way links were built between thematic data and existing and new alphanumeric relational data bases (RDBMS). These links allow loading attribute data into a spatial feature from an RDBMS; loading area, length, and/or location of a feature into an RDBMS; and selecting spatial features based on an RDBMS query list (Green and Collins 1990). These links are effective but constrained by MOSS public domain software. We will be replacing this software as a result of a Bureau wide modernization procurement in the near future.

Completion of the RMPs by 1993 is BLM Oregon's highest priority. The present workload is focusing on completion of the analysis of effects of the various alternative management scenarios including the effects on the forest biodiversity (Nighbert 1991). However, as mentioned in the introduction, this paper will deal with temporary diversions from the RMPs to use the WODDB database to perform special spotted owl project analyses.


DRAIN RESOURCE AREA
SCALE 1: 410000
10.4
0 MILE
10.4
10.4
0
10.4 KILOMETER
N

SPOTTED OWL HABITAT AREAS
**Figure # 1**

## Owl habitat projects

During the period from 1983 to 1987, BLM signed agreements with the State of Oregon to exclude 110 areas from new timber sales to protect the spotted owl. These Spotted Owl Habitat Areas (SOHA's) vary from several small parcels to large blocks of BLM administered lands and were based on known owl pair locations. These SOHA's were considered to be administratively withdrawn from the timber base until the completion of the RMPs, which would address different levels of protection for owls and other wildlife species.

The analysis and graphics developed for the following projects included all BLM lands in western Oregon. However, for readability in this paper, graphics will be presented for only the Drain Resource Area, one of the 20 resource areas in western Oregon. The Drain SOHA's are shown in figure #1. In these figures the closed polygons are Bureau lands and the cross hatched areas are habitat proposals.

### April 1990

The Interagency Scientific Committee report, "A Conservation Strategy for the Northern Spotted Owl" was released. This report, which became known as the ISC report, recommended setting aside major blocks of land for owl habitat. These blocks were designated Habitat Conservation Areas (HCA's). HCA-1's and HCA-2's supported one or more existing nesting pairs. HCA-3's and HCA-4's were owl sites, buffered by either 1/4 mile or 1.5 miles radius (figure #2). HCA-5's were all the other lands, outside HCA-1, 2, 3 & 4's, that didn't meet a "50-11-40 rule". This rule would prohibit timber harvest in a 1/4 township area (3 mile by 3 mile) that did not have 50% of the forest landscape with trees of at least 11 inch Diameter at Breast Height (DBH) and a 40% or more crown closure.

The habitat issue is the amount, size, and distribution of old growth forest now and in the future, habitat considered critical to owl survival. Saving old growth forest is in direct competition with the volume of timber available for harvest and the social economic effects to Oregon, especially small timber towns who's life blood is logging.

In order to address the social and economic issues, it was necessary to compute the Allowable Sale Quantity (ASQ). The ASQ is a sustainable, long term, timber harvest volume. It is a function of the timber stands available for logging and the productivity of these stands. In WODDB, over 65 000 stand polygons are mapped. Each stand can be in any one of three categories: in the timber base, out of the timber base, or partly in the timber base. Actually, most stands are either out of the timber base or partly in the base. For instance, some components of stands such as roads and small islands of non-forest are not suitable for timber growth. Other areas such as stream and visual corridors were withdrawn or constrained from logging based on the 1980s land use allocation. Therefore the ASQ is based on the net land base available, which is the gross land base minus the timber withdrawals for each stand (net acres = gross acres - withdrawn acres).

A spatial problem for the RMPs was how to account for the acres without double counting. A simple example was do we count the acres withdrawn at the intersection of a buffered stream and buffered road as a road withdrawal or a riparian withdrawal. Our solution was to define a hierarchy of spatial withdrawals. At the top of the hierarchy were permanent withdrawals such as non-forest, non-suitable and roads. Lower in the hierarchy required management involvement, such as if a stream is buffered to maximize anadromous fish habitat and roads are buffered to maximize visual resources in the area, how do we account for the acres? These questions need to be addressed by management to determine if we want to credit riparian protection, visual protection, or both. The latter gets tricky since we have over 21 categories of acres in the RMP based on a hierarchy. Therefore, we have established a hierarchy for each of the five management alternatives in the RMPs, based on the management objectives of each alternative.

For the analysis of the ASQ under the ISC report, given current management plans as the datum we need only find current plan net acres, minus ISC acres for each HCA-1,2,3 & 4 by each timber stand. The acres were loaded into the RDBMS Micro-Storms for computation of the 50-11-40 status and development of the strata for input to the ASQ model. The end result was an assessment of the ASQ and social economic impacts, quicker and more accurately than anything we had been able to do pre-WODDB (Green and Collins, 1990). This project we called "Feather", and later became "Feather 1", as we responded to new spotted owl habitat issues.

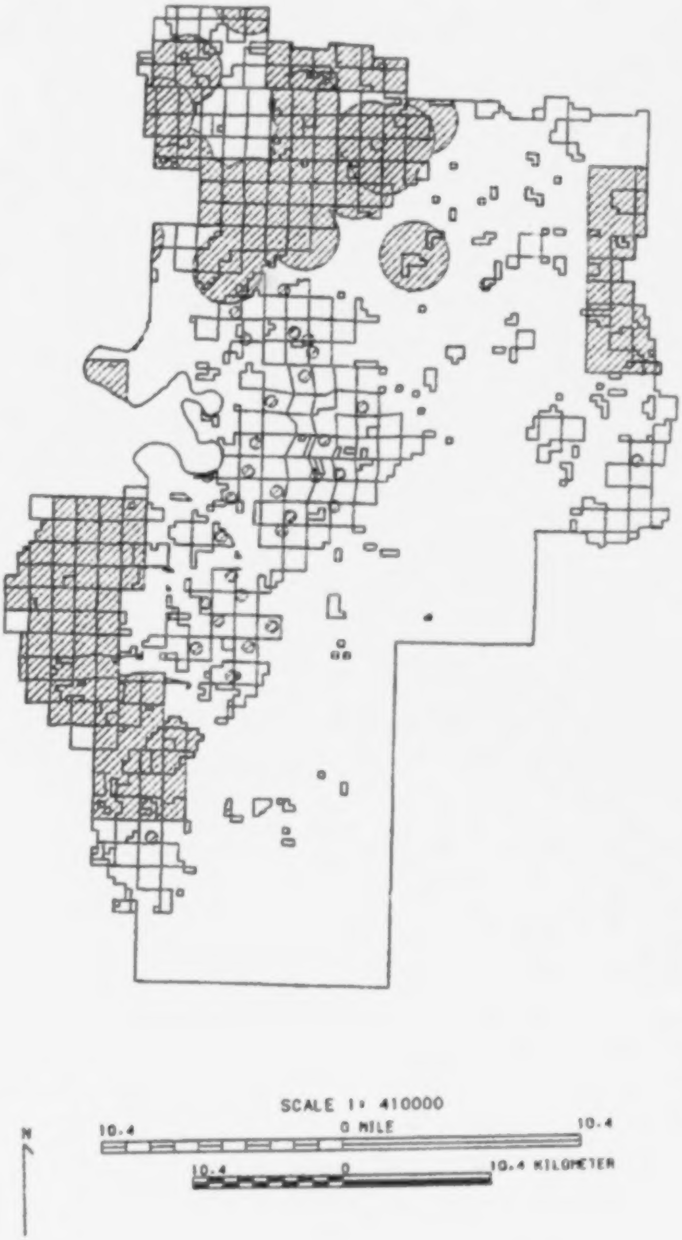
### July 1990, Feather 2

Feather 2 began as the Bureau attempted to define a timber sale strategy given SOHA's, HCA's and the current land use allocation.This was to be an interim strategy until completion of the RMPs in 1993.

In the RMP process we had developed a theme we called potential harvest. This theme provided a plot at 1:12 000 identifying stands potentially available for harvest. District timber sale specialists then developed logical sale units and roads considering streams, roads, topography and their knowledge of the land. These logical sale units were digitized to create the representative timber units (RTU) and representative timber roads (RTR) themes. This was a tedious process but turned out to be time well spent.

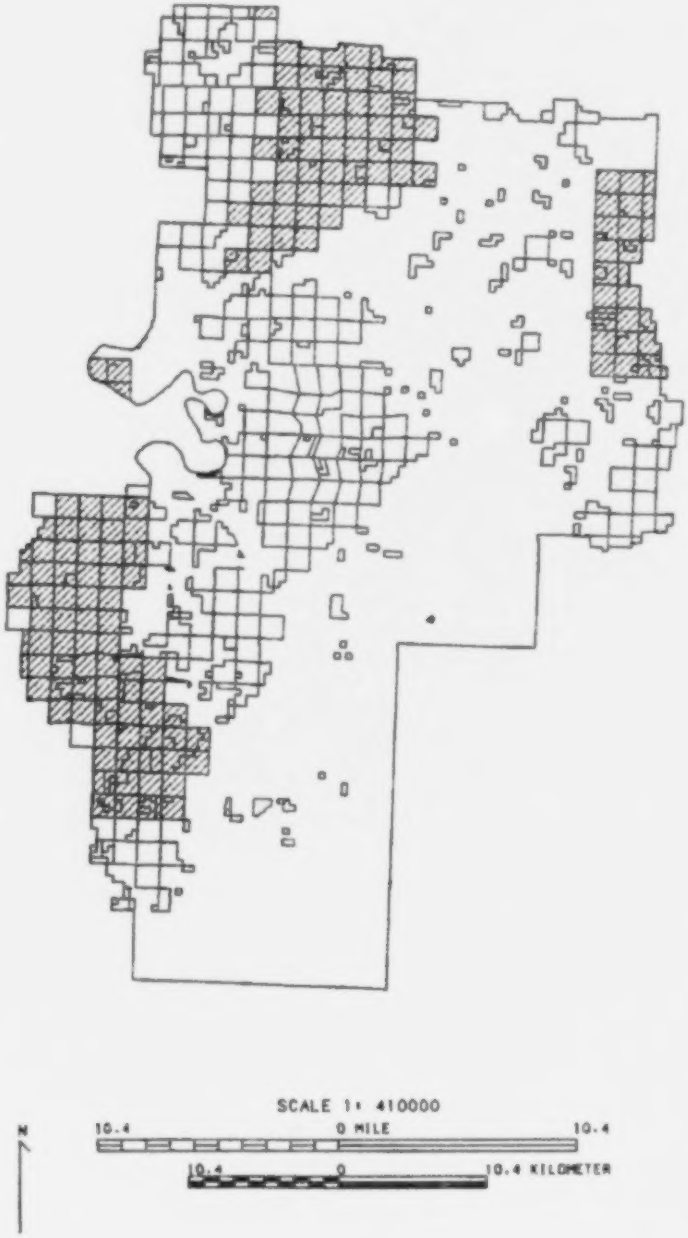
Given we had captured HCA's for Feather 1 and we had completed the RTU theme, we needed to capture SOHA's to give the timber sale staff a clear picture of logical sale units available for harvest. We were still

DRAIN RESOURCE AREA


SCALE 1: 410000
10.4
0 MILE
10.4
10.4
0
10.4 KILOMETER
N

HABITAT CONSERVATION AREAS
**Figure # 2**

DRAIN RESOURCE AREA


SCALE 1: 410000
10.4
0 MILE
10.4
10.4
0
10.4 KILOMETER
N

CRITICAL HABITAT AREAS

Figure # 3

lacking a current (1980s) theme depicting the areas out of the timber base. The district staff had this knowledge from other sources. Therefore, they identified which RTU polygons were selected for the 1990 and 1991 sales. These units were selected from the RTU theme to build the new sales theme. Most of the computational work was completed in the RDBMs since we had completed Feather 1. Plots were sent to Washington, D.C. for use by the Bureau Director in selling his plan, The Jamison Strategy, to Secretary of Interior.

### May 1991, Feather 3

The United States Fish and Wildlife Service (USF&WS) published "Draft Proposal for Northern Spotted Owl Critical Habitat" in the Federal Register. Actually the critical habitat areas (CHA's) were published in two forms: a legal description, and an AutoCAD drawing. Our first challenge was to capture this theme (figure #3).

Since the CHA descriptions didn't include any meets and bounds land descriptions, we selected the polygons from our ownership theme based on the legal description. We did have to resolve errors in the legal descriptions which put some CHA's in the Pacific Ocean. The question the BLM needed to address again was the effects upon the ASQ and the resulting social and economic impacts.

In our spare time since Feather 2, we had been working on the RMP. One of the processes nearly complete was the Alternative B out acres. This gave us the acres of each stand that would be out of the timber base in the 1980s land use allocation. Since this data was in the RDBMs and stands were dissected by legal description, we were able to use the legal description as the query in the RDBMS to compute the remaining timber inventory and hence the ASQ. GIS was used mostly for graphics.

### June 1991, Feather 4

Congressional sub-committees established a "Congressional Advisory Panel on Pacific North West Old Growth Forests". The challenge of the advisory panel, know as the "Gang of Four", was to identify late successional forest for interim protection until permanent solutions were found. In addition, they were to define an interim management strategy for the remaining forest and the resulting ASQ for any alternatives. BLM, Forest Service and USF&WS staffs from Oregon, Washington and California met in Portland with 1/2 inch to the mile base maps.

We had been looking at various scenarios related to old growth forest and different definitions of late successional stands. Therefore, we were able to provide staff with rescaled plots defining 200 year plus stands (figure #4), 120 year plus stands with 40% plus crown closure (figure #5), HCA-1,2,3 & 4's (figure #2), BLM ownership, spotted owl center of activities and registration grids. Basically we provided plots from digital data for hand analysis to be captured for digital displays and graphics! The study was completed and published. The results were made available to the spotted owl recovery team discussed as Feather 8.

### July 1991, Feather 5

This project was a self inflicted wound, as the BLM Director wanted us to produce a 40 year age class map for all stands for the years 1990, 2000, 2020, 2030, 2040 and 2050. This would assume no harvest for the next 60 years. This was used to assist in the understanding that, due to the BLM's age class distribution and land fragmentation pattern, a no harvest scenario would not provide large blocks of old growth forest. These no harvest themes were created by growing stands from the birthdate of the stand. Given the data structure of WODDB, which is limited by the current software, these six plots required 64 826 GIS selects (6 plots, 23 age classes, 477 townships).

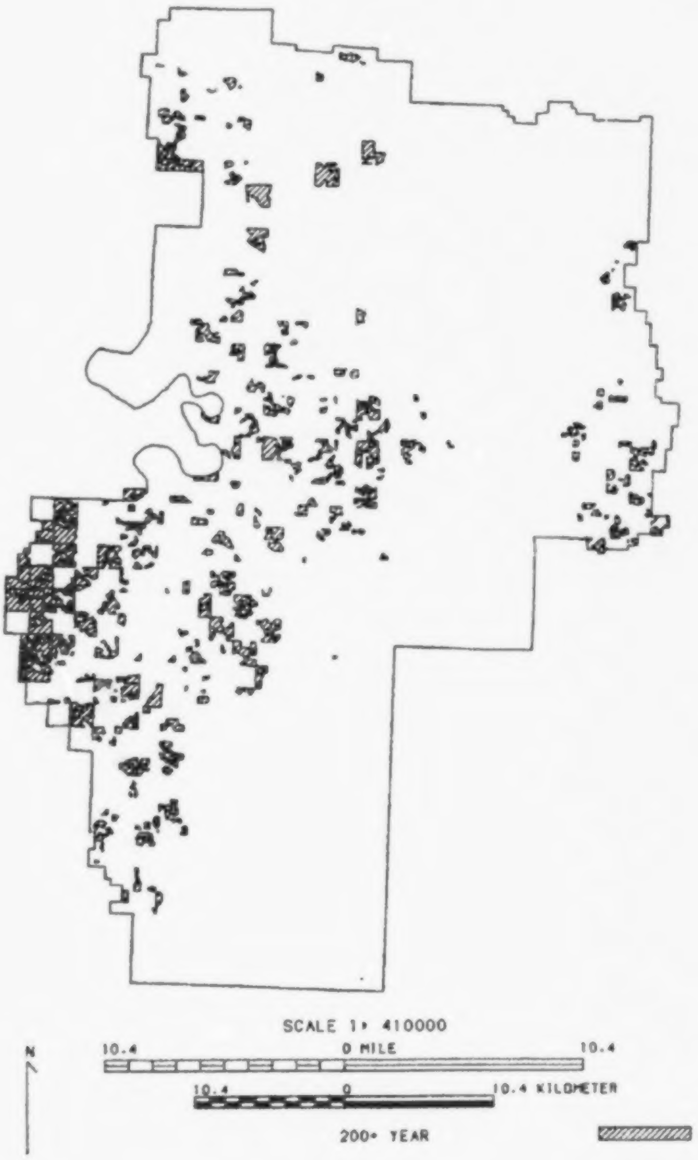
### August 1991, Feather 6

The USF&WS published a "Revised Proposal for Northern Spotted Owl Habitat" (figure #6). This was a revision of Feather 3's CHA's (figure #3). The critical habitat units (CHU's) were published in the Federal Register with legal descriptions and AutoCAD drawings. This time the USF&WS used a GIS to define CHU's so a digital version was available.

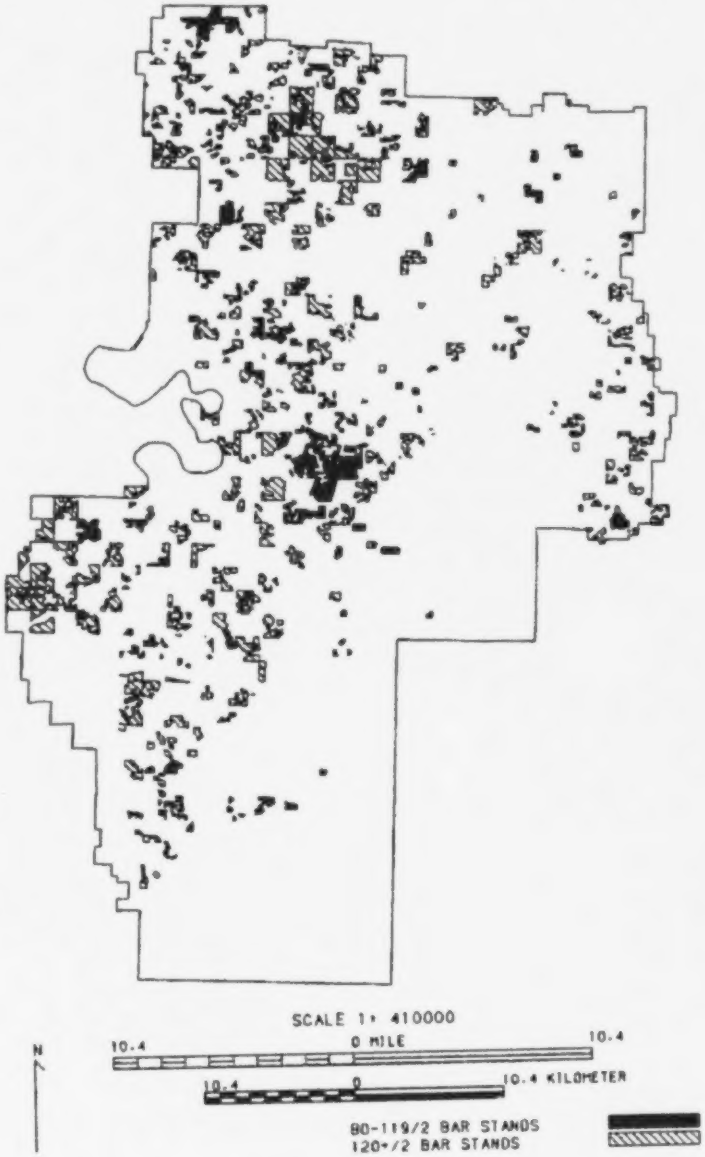
However, one of the early decisions we made with the WODDB project was to use an updated public land survey system (PLSS) grid (Crouser 1990). What this means is our 477 WODDB township lines are more refined than those shown on United States Geological Survey (USGS) 1:24 000 quad maps. Therefore, whenever we are dealing with a legal description we need to capture (snap to our PLSS grid) the theme, even if it's been digitized at a compatible scale by another agency. So we compiled this theme from the ownership theme as we did in Feather 3. Again the challenge was to find the effects of these CHU's on the ASQ and the social economic impacts. The analysis of the ASQ was completed in the RDBMS, with GIS used mostly for the graphics. New CHU's were intersected with timber sales and used the existing Timber Sale Information System RDBMS to find the volume, value and status (sold, contract awarded, pending) of sales inside the new CHU's.

### October 1991, Feather 7

GIS support for the Spotted Owl Exemption Sales Committee (popularly known as the God Squad) began in late October, 1991. At issue here are 44 of our 1990 timber sales that failed to pass the consultation process, required by the Endangered Species Act, with USF&WS. The Secretary of Interior has invoked this committee to determine whether an exemption to the Act should be


SCALE 1: 410000
N
10.4
0 MILE
10.4
10.4
0
10.4 KILOMETER
200+ YEAR

DRAIN RA OWL HABITAT 1
**Figure # 4**

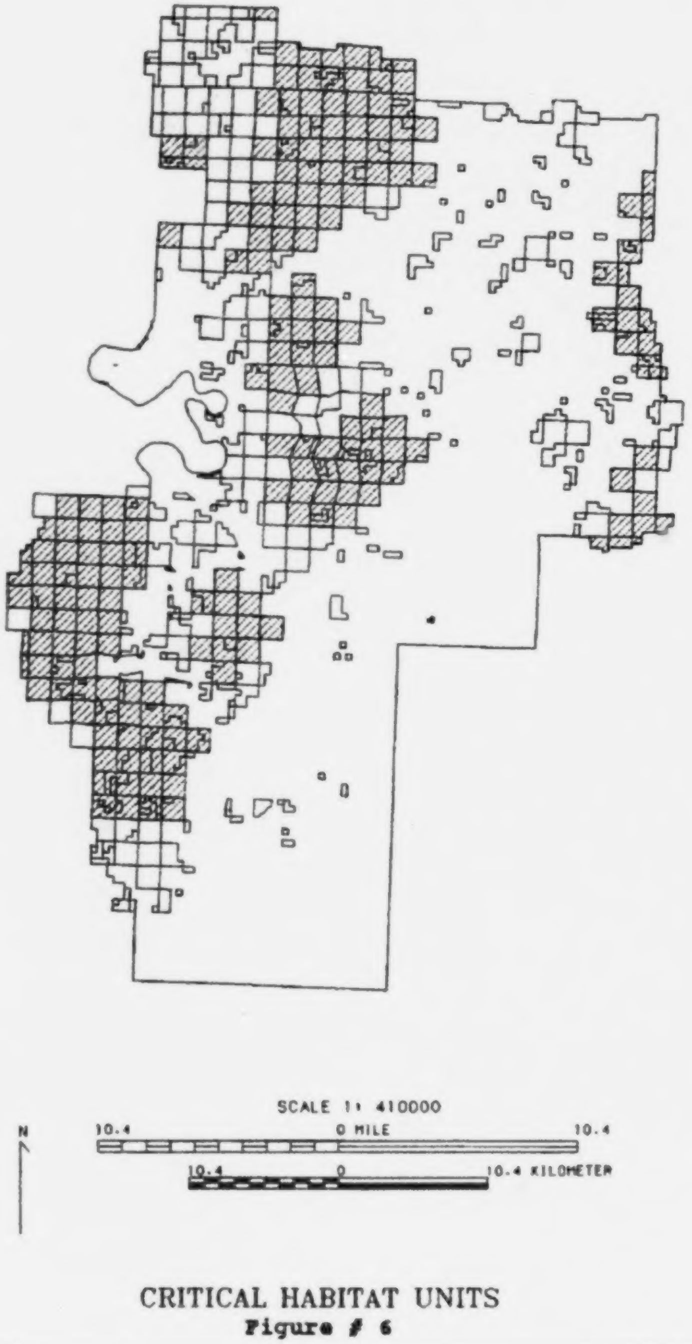


SCALE 1: 410000
10.4
0 MILE
10.4
10.4
0
10.4 KILOMETER
N
80-119/2 BAR STANDS ■
120+/2 BAR STANDS ▨

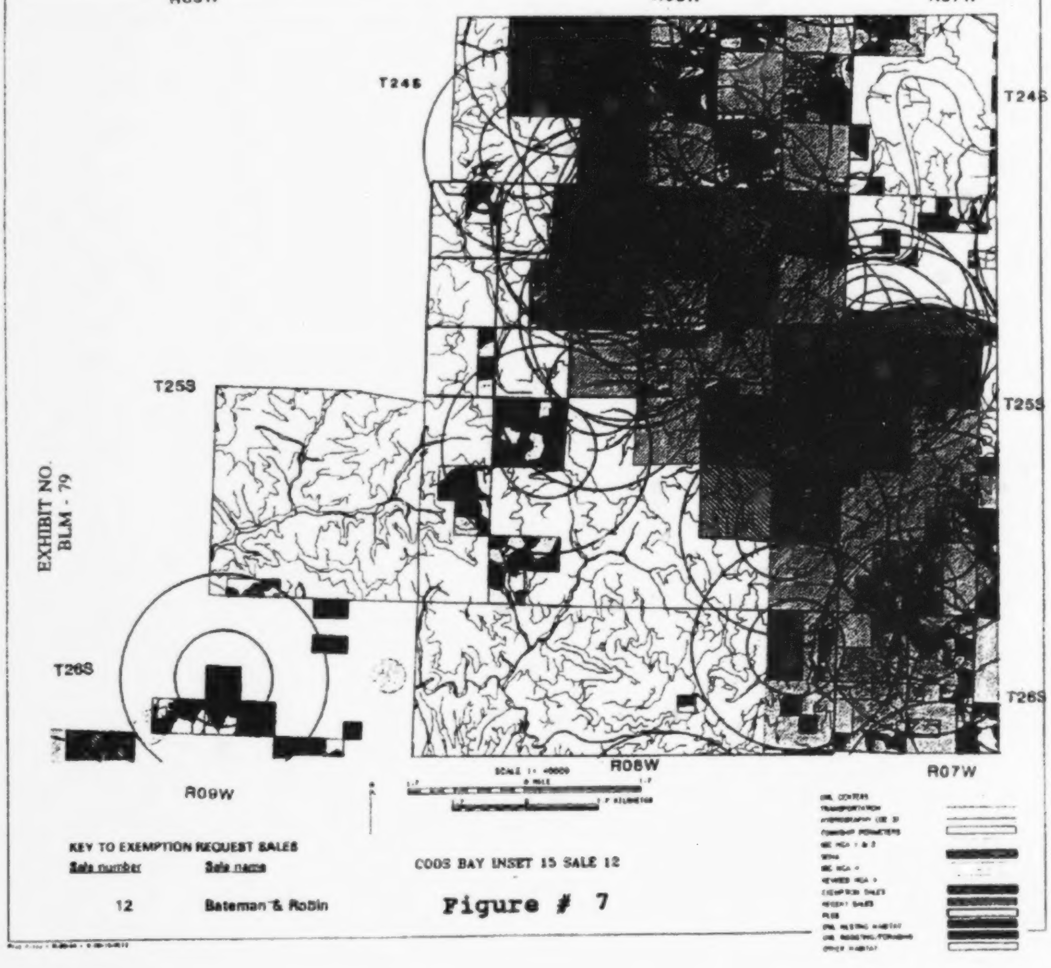
DRAIN RA OWL HABITAT 2

**Figure # 5**

DRAIN RESOURCE AREA


SCALE 1: 410000
10.4
0 MILE
10.4
10.4
0
10.4 KILOMETER
N

CRITICAL HABITAT UNITS
**Figure # 6**

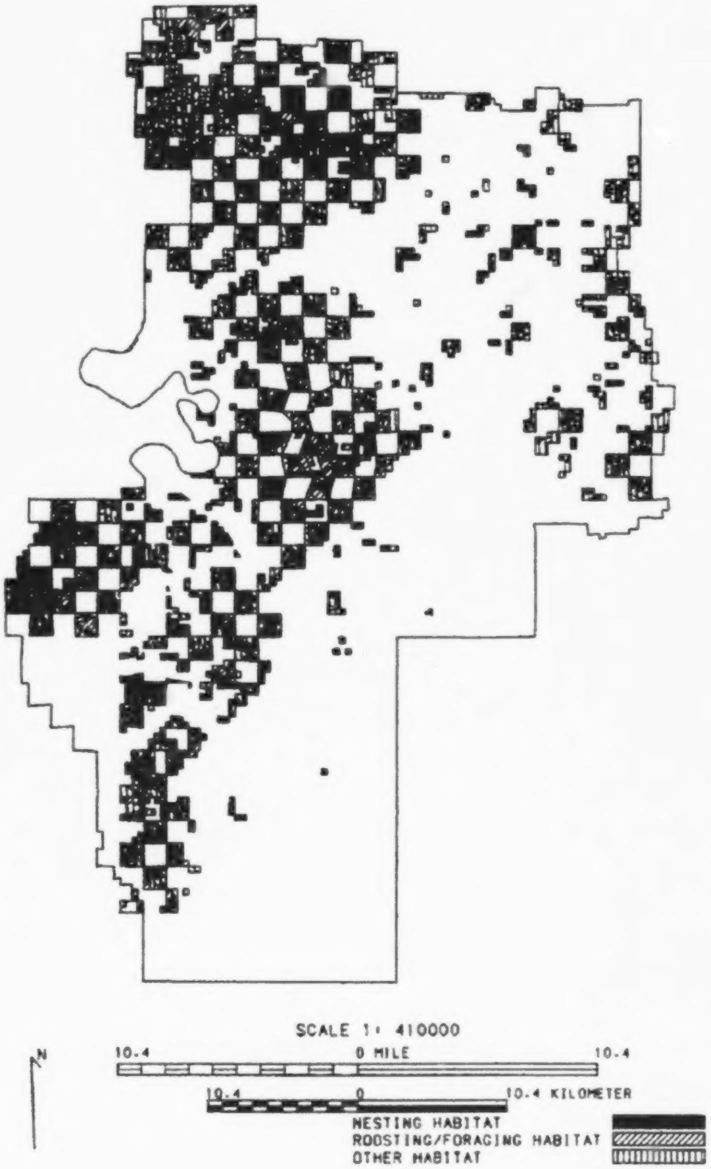

R09W
R08W
R07W
T24S
T24S
T25S
T25S
T26S
T26S
R09W
R08W
R07W
EXHIBIT NO.
BLM - 79
SCALE 1 : 48000
OWL CENTERS
TRANSPORTATION
TOWNSHIP PERIMETERS
EXEMPTION SALES
RECENT SALES
OWL NESTING HABITAT
OWL NESTING/FORAGING
OTHER HABITAT

KEY TO EXEMPTION REQUEST SALES

| Sale number | Sale name |
|---|---|
| 12 | Bateman & Robin |

COOS BAY INSET 15 SALE 12

Figure # 7


SCALE 1: 410000
10.4
0 MILE
10.4
10.4
0
10.4 KILOMETER
N
NESTING HABITAT
ROOSTING/FORAGING HABITAT
OTHER HABITAT

DRAIN RA OWL HABITAT 3

Figure # 8

granted for these sales. In order to present a credible set of visual exhibits, the 44 exemption sales boundaries were captured. Prior to this we had only the RTU boundaries of these sales in the WODDB, which was a rough approximation of the sale boundaries.

The Oregon Governor's Office identified 81 communities severely affected by the lack of timber. These were captured from 1:100 000 scale USGS quads. Other themes were created using the Spotted Owl locations we had and buffering them by different widths based on their locations within certain zones, again a theme we already had in the system.

We initially provided the lawyers a plot showing location of the 44 sales. They asked for a better spatial perspective showing the sales, main state and federal highways and severely affected timber dependent communities.

Support for the witnesses required more analysis and special representations of the data. We provided plots at 1:40 000 scale that put the sales in perspective with the lands around them. One of the repeated questions from the USF&WS is why we did not pick other locations for these sales. Plots were created that merged BLM ownership, the lands taken out of base by the SOHA's, the Habitat Conservation Areas (HCA-1, 2, 3 & 4's) established by the Interagency Scientific Committee, the additional lands identified under the ISC report for dispersal habitat (50-11-40 rule), and the recently raised requirement from the USF&WS Biological Opinion that we do not harvest timber within 1.2 to 1.5 miles of any owl nest based on geographic province (figure #7).

These plots show little land left available for timber harvest. Beside the 1:500 000 scale original plots that will be used in testimony, we also provided copies of the plots as 11x17" versions that will be reproduced and made part of the permanent record.

In order to fully examine the affect on BLM acreage available for timber harvest, we also intersected the expanded owl buffers with BLM defined habitat (figure #8) to get the total number of acres in BLM ownership and in each habitat type (category 1 = Nesting; category 2 = Roosting and Foraging; category 3 = Other). This exercise was based on BLM biologists attributing every timber stand by the above categories, rather than a standard query such as 200 plus year old stands.

To show the relative timber age of the adjacent non-BLM lands we provided plots of BLM ownership and the Exemption sales at a scale of 2 inches to the mile to fit 1991 aerial photography.

**December/January 1992, Feather 8**

The plot continues. The Spotted Owl Recovery Team is due to release their report, which outlines management of critical owl habitat. The Final Critical Habitat designations are due to be published. One hopes that at some point there will be an agreement on what is critical habitat, where is it and how it should be managed.

## Summary

Since April of 1990 we have computed the ASQ for western Oregon nine variations of spotted owl habitat proposals. Prior to WODDB this would have involved dot grid compilation of acres on 65 000 timber stands, nine times, dividing each stand into acres available and not-available for timber harvest based on buffered owl sites, buffered roads, buffered streams, SOHA's, HCA', CHA's, CHU's and land use allocations. This totals around 3.5 million polygons. To say it couldn't be accomplished by hand would not be true, as BLM had to do it! However, these projects couldn't have been accomplished by hand in the time frame allowed or to the degree of accuracy required without WODDB. Our investment in WODDB began to pay off in April 1990, and continues today.

I have spent most of my time discussing how the Bureau is responding to different physical and spatial definitions of spotted owl habitat. The paper by Dippon *et. al.* 1992, discusses the linking to the McKelvey model. This model, HCA's, CHA's, CHU's, our biologist's habitat calls, and the results of the analysis of the five RMP alternatives, including a forest biodiversity analysis (Nighbert 1991), will be utilized in developing the last stage of our RMP process, the preferred alternative.

## References

Crouser, Lisa. "Developing the Foundation: A Geographic Coordinate Data Base for Forest Resource Planning". Vancouver, British Columbia, GIS'90 Symposium. pp. 389-394.

Dippon, Duane R. "Building the Western Oregon Digital Database for the 1990's Resource Management Planning Effort" by Duane R. Dippon, Pat Green and Don Pearson. Orlando, Florida, GIS/LIS '89 Proceedings, November 1989. pp. 276-285.

Dippon, Duane. "Linking a Spatially Oriented Northern Spotted Owl Population Dynamics Model with the Western Oregon Digital Database" by Duane Dippon, Chris Cadwell, Jeff Nighbert, and Kevin McKelvey. Vancouver, British Columbia, GIS'92 Symposium.

Green, Pat. "Maiden Voyage of WODDB" by Pat Green and Dana Collins. Seventh National MOSS Users Conference, November 1990, Portland, Oregon.

Nighbert, Jeffery S. "The Application of ADS/ORACLE/MAPS to Forest Biodiversity in Western Oregon." Eighth National MOSS Users Conference, December 1991, Bozeman, Montana.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Modelling Dall Sheep habitat in the northern Yukon

**Manon Desforges**
**Lauren Crooks**
GIS/Remote Sensing Unit

**Jean Carey**
**Val Loewen**

Fish and Wildlife Branch
Yukon Renewable Resources
PO Box 2703
Whitehorse, YT, Canada Y1A 2C6

## Abstract

*GIS is becoming a powerful tool for mapping, planning and research in the Yukon. The Yukon Department of Renewable Resources is using GIS technology to help determine the relationships between wildlife distributions and habitat parameters.*

*Winter ranges, lambing areas, rutting grounds, migration corridors and mineral licks are all key habitat for Dall Sheep. This study was limited to the analysis of lambing areas. In the first phase, spring sheep locations were related to slope, aspect, proximity to escape terrain and proximity to winter range. With the modelling capabilities of the GIS we then predicted, in the second phase, lambing sites for a test region using the relationships identified in the first phase. The accuracy of the model was finally verified by overlaying surveyed lambing locations in this test area with the predicted lambing areas.*

*The results could be used to assess impacts of development scenarios on wildlife and assist in land use decisions in the northern Yukon. Future work will continue to include other wildlife species in various regions of the Yukon.*

# C3 GIS in information management

**Mapping and establishing GIS for tropical rain forest in Indonesia**

Herman Hidyat, PT Adikarto Printindo, Jakarta, Indonesia

**Reliability of GIS for environmental assessement in data poor regions**

Kevin Barthel, University of Wisconsin-Madison, WI, USA

**Using GIS and image processing to prioritize cumulative effects analysis**

Steven Bernath et al., Washington State Department of Natural Resources, Olympia, WA, USA

**An examination of the role of GIS in the regulatory process: the Oslo project as an example**

James Case et al., CASEBiomanagement, Calgary, AB, Canada

**Monitoring and modelling transboundary air quality using a GIS**

Anne Lucas, IBM - Bergen Scientific Centre, Bergen Norway

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Mapping and establishing GIS for tropical rain forest in Indonesia

**Herman Hidayat**

General Manager PT. Adikarto Printindo
Geding IKPT Lantal-2
JL Bendungan Hilir Raya No. 50
PT. Mapindo Parama
Jakarta Indonesia
Tel: 570-3168 Fax: 570-3170

## Abstract

*The total area of Indonesia is about 780 million hectares and consists of sea (587 million hectares) and land (193 million hectares). Seventy five percent of the land is covered by forest and fifty percent of that forest is so-called* Production Forest. *Wood products contribute up to 30% of the national income from exports.*

*In order to achieve sustainable development in the forest industry and maintain the forest as a renewable natural resource, PT. Mapindo Parama is now conducting a survey and mapping of Production Forest throughout Indonesia, a total of 70 million hectares. Forestry spatial and descriptive data are being collected and later on the* Forest Resource Information System *will be established. All activities are planned to be finished within 5 years.*

*In establishing the database, PT. Mapindo Parama employs not only aerial photos at a scale of 1:20 000 but also SPOT, Landsat Thematic Mapper and Synthetic Aperture Radar Imageries. All data are being compiled and stored in digital form.*

*This is a report on the mapping and establishment of the GIS Project for the richest tropical rain forest in the world.*

## Introduction

Indonesia has the third largest Tropical Rain Forest in the world, but probably the richest one. The Government of Indonesia (GOI) has determined the Forest Land uses as shown on the next page.

The *Protection Forest, National Forest and Production Forest* are determined as a *Fixed Forest* while the *Converted Production Forest* is an area of forest which can be converted into other land uses such as transmigration settlements, agriplantations, etc.

The *Production Forest* area spreads out in several islands thoughout Indonesia: 15 million hectares in Sumatera, 30 million hectares in Kalimantan, 5 million hectares in Sulawesi, 3 million hectares in Maluku and 11 million hectares in Irian Jaya. The distance between Sumatera and Irian Jaya is almost the same with that between the west coast and east coast of the USA. Currently the GOI has given the right to manage 60 million hectares of production forest area, usually called *forest concession rights*, to 578 Concessionaires.

Those Forest Concessionaires established the Association of Forest Concessionaires which has worked hand in hand with the GOI, in this case, Ministry of Forestry, to develop the best

| | |
|---|---|
| • Protection Forest | 30 million hectares |
| • National Forest/National Reserve | 19 million hectares |
| • Production Forest | 64 million hectares |
| • Converted Production Forest | 30 million hectares |
| | 143 million hectares |

systems for managing the Indonesian Tropical Rain Forest. Both parties realize that achieving sustainable development in the forest industry and maintaining the forest as productive contributes significantly to the Indonesian economy currently and in the future.

To support the above policy, the Ministry of Forestry and Association of Forest Concessionaires have begun to conduct a forest inventory, monitor and, later on, to establish a Forest Resource Information Systems.

## Contribution of forest products to the national revenue from export

Before oil prices crashed in 1986, Indonesia was famous as one of the oil producing countries. Almost seventy percent of its national revenue from export was oil money. In 1985 non-oil exports contributed only 32% to the national revenue. But gradually, these non-oil exports have played a more and more important role. Currently oil money contributes only 40% while non-oil products contribute 60% to national revenue.

One of the components of non-oil products is wood product. While the income from wood product was US$ 1.3 billion in 1985, it tripled in 1990 and is projected to be US$ 8-10 billion every year in the next five years.

The forest industry absorbs almost 3.5 million people in various related industries from logging companies to woodworking factories. That means that the forest industry in Indonesia is a very labour-intensive industry. If one family consists of 4 people (husband, wife and 2 children), it means this forest industry sustains 14 million peoples' lives.

Realizing this situation, the Indonesian Forest Community has committed to achieve sustainable development in the forest industry and maintain the forest as a renewable natural resource.

## Sustainable development in forest management

Forest Inventory plays an important role in obtaining better information regarding forest resources, such as merchantable, and lesser known species, their diameter distribution, topography, soil conditions and climate. To attain information on growth, inventories on primary forests and residual stands will be undertaken. Residual stand inventories will be done directly after logging operation, and will be repeated five year thereafter and one year prior to harvesting. Based on the characteristics and nature of each forest, which vary greatly from place to place and from type to type, the policy regarding the size of management unit areas, age cycle, cutting cycle and allowable annual cut can be formulated.

In order to maintain sustainable forest management, the MOF employes a progressive sustained yield principle. That is why the MOF imposes 3 silviculture systems which are:

1. TPTI which is Indonesian Selective Cutting and Replanting System, a version of the shelterwood system. The TPTI System is based on the amount and diameter of nucleus trees per hectare. The minimum diameter for allowable cutting tree is 50 cm, while residual stands should have a diameter of more than 20 cm and minimum 25 trees per hectare. The cutting cycle is 35 years.
2. THPA is a clear felling with natural regeneration.
3. THPB is a clear felling with artificial regeneration. One example of this system is HTI (Industrial Timber Plantation Forest).

Each system above has its specification criteria. Which system to be employed in a production forest area will be determined by the MOF after evaluating the inventory survey results carried out before. Currently, most of the Forest Concessionaires are using the TPTI system, but the MOF has also encouraged some of them to do HTI.

## Forest inventory and database

Good forest management is impossible without good inventory data. In this case, forest inventory is considered to be synonymous with timber estimates. That is why forest inventory is an attempt to describe the quantity and quality of forest trees, especially the merchantable tree species, and many of the characteristics of the land area upon which the trees are growing. So, we can categorize this forest inventory as timber-oriented inventory.

The purpose of forest inventory is to establish a database of the entire Production Forest throughout Indonesia. In the first stage, some activities *(following)* have been done:

a. Aerial Photography at a scale of 1:20 000
b. Photogrammetric Mapping to produce form line map at a scale of 1:25 000.
c. Photo interpretation to produce Vegetation Map at a scale of 1:25 000.

### Base maps and ground control points

The National Agency for Survey and Mapping - BAKOSURTANAL - is a government organization which is responsible for establishing ground control points networking and base maps in Indonesia. Unfortunately, they have not finished the work yet. The area in which Base Maps 1:50 000 and Ground Control Points are already finished are Java, Bali and Sumatera Islands and some areas along the coast in Kalimantan and Sulawesi. For the rest there are only JOG Maps at a scale of 1:250,000 available, which were published by the US Army Map Services.

Facing this problem, the Ground Control Points network will be established in the early stages by employing the Global Positioning Systems method for the area which have no base maps or Ground Control Points.

Improper base maps in certain areas also cause navigational problems in taking aerial photographs, that then result in abundant stereo models of photograph. In average one stereomodel 1:20 000 contains 300 hectares only.

### Aerial Photography

Panchromatic B/W aerial photography at a scale of 1:20 000 is being done throughout Indonesia to cover the entire Production Forest of 64 million hectares. The company now operates 10 aircraft equipped with aerial survey cameras with focal length of 15.2 cm. The production rate of photography is about 2 million hectares per month. This seems to be a very low production rate, but this is actually quite a reasonable rate since Indonesia lies in the equator. So, one of the biggest problems is cloud cover. Every month approximately 6500 stereo models are produced.

### Photogrammetric mapping

In order to produce Form-Line Maps with the scale of 1:25 000 some analytical and analog stereoplotters are used. Since all data are processed digitally, the analog stereoplotters are converted into semi-analytical ones. Using 5 Zeiss Planicomp and 6 Wild B8 Stereoplotters, almost 40% of the work is done in the company, while the rest of the job is to be subcontracted to other Indonesian mapping companies.

The size of the map is 55cm x 55cm; that means one sheet of map contains approximately 18.900 Ha. Maps editing, joining and sheet creation are done digitally. All systems are working in AutoCad environment-PC Version.

### Vegetation mapping

Photo-interpretation is done in order to get information on vegetative cover type, crown density, stand height, Crown diameter and to derive categories such as volume classes. Stratification of the forest is done on aerial photographs with a forest class map as follows:

- Crown closure density (C)
    - C1 : 10 - 40%
    - C2 : 41 - 70%
    - C3 : 71 - 100%
- Stand height (H)
    - H1 : 10 - 20 m
    - H2 : 21 - 30 m
    - H3 : > 30 m
- Crown diameter (D)
    - D1 : < 10 m
    - D2 : 10 - 20 m
    - D3 : > 20 m

Forest photo interpretation is concerned with recognition of forest cover types, the deliniation of areas, the measurement of trees and stand and the estimation of timber volume afterward by combining them with ground sampling results using a regression analysis method.

The volume of forest stands can be estimated on aerial photographs to the extent that it is correlated with tree heights, Crown width, and Crown closure. Based on ground sampling, the stand aerial volume table predicts the average stand volume expected for a given stand height and crown closure. Crown width is sometimes also taken into consideration.

Using 6 Zeiss Stereocord and manual mirror steroscopes almost 40% of the work is done in the company.

### Spot and Landsat TM

For the areas which have not been covered yet by the aerial photos, SPOT and Landsat TM images are being used, especially those which were taken after 1988. As aerial photographs, the spaceborne images which are cloud free are difficult to get.

Based on the images, a general specification of forest cover type can be determined. This technique is usually applied to determine whether virgin or primary forest areas or Industrial Timber Plantation Forest Areas are to be established. Interpretations of the images are done in two ways, i.e. manually and digitally, through the image processing system method.

The company has been using 156 scenes of SPOT images covering approximately 20 million hectares owned by 344 Concessionaires and 7 scenes of Landsat TM image covering approximately 10 million hectares.

**Synthetic aperture radar (SAR)**

It has historically been very difficult to obtain airborne aerial photographs or optical satellite imagery of almost 20 million hectares of production forest areas as they have been obscured by cloud. Toward the area mentioned above and integrated multi-system approach of substitution Remote Sensing Technologies will be applied.

The company and Ministry of Forestry are now conducting a Pilot Study on the application of SAR for forest inventory and management in Tropical Rain Forest Areas, especially over mountainous or generally hilly areas.

The overall concept for the Pilot Study Program is as follows: the first phase of the program would be to collect Synthetic Aperture Radar (SAR) image data over the study area in Kalimantan to provide guaranteed, high-resolution, cloud-free imagery. The high resolution SAR Mode provides resolution of 6 x 6 m using a swath of 23 km. SAR data collection is carried out in side-looking mode with a nominal 70% overlap between adjacent swaths. In the second phase of the program, helicopter mounted low-altitude photography (heliboom) would be collected to provide sample plot information within this study areas. The third phase of the program would possibly be collecting laser profile data in grid over the study area, including the sample plot areas.

The SAR data are used to provide complete broad area coverage, to generate topographic contour and forest cover types information, and to select sites for heliboom photography sampling. The heliboom photography will be used to provide the density and crown diameter information as well as species mix data. The laser profile information will provide tree height information and be used to tightly register the topographic contours. The sample plot and field sample data will be extrapolated to cover the entire study area using the interpreted SAR data as a guide. Further field visits will be make to the study areas to assess the accuracy of the extrapolation. Recommendations will be made to optimize the operational use of the technologies demonstrated.

## Monitoring systems

In general, reinventory of forest will be done every 5 years to update the established data base and for monitoring purposes. For general and broad view spaceborne images (SPOT, Landsat TM, Radarsat, etc.) will be employed, but toward specific areas which are identified from spaceborne images are needed to get more detailed data will be photographed using fixed wing aircraft and heliboom. An integrated image processing system will be established.

Annual field reports provided by Forest Concessionaires will also be used as a tool for monitoring and updating of data base purposes.

## Forest resource information systems support forest management

Geographic Information Systems will be established in the company to serve the Association of Forest Concessionaires, Forest Concessionaires Companies and also the National Forest Inventory Program done by the Ministry of Forestry.

The Association and Ministry of Forestry need the information for strategic policy and Forest Management purposes, while Forest concessionaires companies need the information for operational and forest management purposes. A system design of GIS and prototype was finished last month. It is planned that the GIS activity will be started in July this year.

## Conclusion

Establishment of a database of Forest Inventory and Forest Resources Information Systems (FRIS) in Indonesia as the Third Largest Tropical Rain Forest in the world needs big efforts and strong commitments to accomplish the task. Mapping, Forest Inventory data and FRIS are very important tools to support management of forest as a renewable natural resource.

**The physical well-being of people everywhere depends upon the way they use their natural resources** *B. Husch — FAO*

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Reliability of geographic information systems for environmental assessment in data poor regions

**Kevin M. Barthel**

University of Wisconsin-Madison
Land Information and Computer Graphics Facility
Institute for Environmental Studies
Land Resources Graduate Program
25 Agricultural Hall
Madison, WI
USA 53706

**Glenn Morgan**

The World Bank
Asia Technical Agriculture Department

**Stephen J. Ventura**

University of Wisconsin-Madison
Department of Soil Science and
Institute for Environmental Studies

## Abstract

*The intent of this paper is to advocate the use of GIS in environmental assessments with cognizance of limitations inherent in GIS processes and data. The implication of these limitations are discussed in the context of environmental assessment and natural resource planning in 'data poor', developing countries.*

*GIS analyses are constrained primarily by limitations inherent in the source material used to create the data base, such as base maps and imagery, and to a lesser degree by the errors introduced in the computerization of these data. Limitations in data quality degrade the reliability of GIS results, and consequently restrict their value in the decision-making process. The analytic power of GIS can also be used inappropriately. Like statistics, GIS is based on core concepts and assumptions that, when ignored, can lead to erroneous or misleading results. These issues are especially significant when GIS is used to study areas where data simply do not exist or where existing data are of poor quality. Such is frequently the case in many developing countries.*

*The authors conclude that GIS can fulfill an immediate need in the environmental assessment of development projects. However, in order for project decision-makers to use GIS results to make better assessments, and more financially and environmentally certain decisions regarding project areas, they must have an understanding of the constraints inherent in the information provided by GIS. Because data are often incomplete, out-of-date or inaccurate in data poor regions, GIS should be used to encourage local and international interaction in the project planning process, and to determine problem areas for further in depth study, rather than as a definitive planning tool.*

## Introduction

The importance of striking a balance between economic growth and the impact of this growth on the natural resource base and natural environment of a country, is becoming increasingly apparent. However, in many developing countries seeking economic development, natural resource reserves constitute a significant, if not the major component of national capital endowments (Bromley, 1986) needed to promote economic development. Concern regarding the use of these resources, combined with the possible degradation of the natural environment has lead to increased public and environmental organization scrutiny of development projects. This scrutiny has also increased awareness by lender and borrowing countries as to the local and global environmental effects of development projects.

Natural resource based development projects that appear economically feasible, can be detrimental to the long term sustainable development of the area without informed decisions based on reliable information[1]. In the case of one multinational development lending agency, the World Bank, the requirement of an environmental assessment (EA) for any project having the potential of causing adverse environmental impacts has been initiated. Results from the EA seek to provide project decision-makers with information on ways of improving projects environmentally, and minimizing or compensating for adverse impacts early in the project planning process (World Bank Operational Directive 4.00).

While the EA requirement is relevant, and it is apparent that EA's will become increasingly important in supporting development project decisions, to be effective, each EA will require diverse information on the natural resource base as well as the existing ecology and infrastructure of the project area. Responding to the requirement for EA's, project managers are now realizing the need for information and the necessity of new tools to help in gathering it, analyzing it, and identifying actual impacts in practice.

## Need for geographic information to support development projects

Economic development policy has frequently been formulated without adequate information pertaining to the countries natural resource base, the existing environmental condition or the environmental impact of proposed development projects. Typically, physical barriers, cultural differences and political impediments combine with the 'datapoor'[2] conditions commonly found in developing countries to make performing an EA a costly, time-consuming and complicated task. As a result of the EA requirement and the difficulty in performing them, there is concern that without new approaches, project managers and decision-makers may rely on existing, possibly out-of-date assessments and unreliable information in their decision process.

Land information is particularly important to developing nations in order to prevent wastage of scarce resources (Dale and McLaughlin, 1988), and to provide a basis for rationally derived, environmentally aware, economic development policy. In many developing countries, accurate land information is not readily available, and that which is available is commonly incomplete or out of date (Morgan, 1990). This deficiency has become apparent, and is well recognized throughout the development field. According to a 1986 internal World Bank study, development projects requiring large and small scale map data, accounting for nearly $800 Million (US), were delayed due to lack of adequate land records and land information. This figure will undoubtedly rise as more lending is targeted for environmentally oriented programs, (Morgan,1989).

Several geographic data collection programs are now established. Examples of these are, the African Remote Sensing Program (Hassan, et.al., 1991), and the Famine Early Warning System (USAID) in Africa, the Regional Remote Sensing Program in Asia and the Pacific funded by the United Nations Development Program (UNDP) (Changhchui, et.al., 1991, Brooner, 1988), and the UNDP Global Environment Monitoring System (GEMS). Many developing countries, however, have either no collection program, or no effective strategy of managing and analyzing these data for use in an integrated development plan. An integrated approach seeking to coordinate individual country efforts and needs, with development agency data acquisitions has been proposed by the World Bank (Hassan, et.al., 1991).

The importance of natural resource data and lack thereof, is recognized throughout the development community. A 1988 United States Agency for International Development (USAID) report states, the principal need in the developing countries is for information which will support the production of food, specifically natural resource information. The report expressed the need for increased use of remote sensing and geographic information systems (GIS) technologies to provide this information.

USAID's involvement with GIS is similar to the World Bank's approach. The intent of these agencies is to encourage the use of GIS in projects where the benefits are compelling. This is done by preparing and disseminating reports describing GIS applications in development projects, determining and reporting on the costs/benefit of GIS, and providing technical support to project planners in preparing technical specifications for the use of GIS/Remote Sensing/Surveying and Mapping technology in development projects (Brooner, et.al. 1988). (Author's note; cost/benefit reports, which significantly influence the use and value of GIS, are lacking. A research proposal examining this issue has been proposed to the World Bank (Jagannathan, et.al., 1990)).

Currently there is no commonly accepted method of analysis used in the EA process to alert policy makers to the environmental consequences of their actions. However, with much of the potential impact of development projects being spatial in nature, the geographic framework and spatial data handing capabilities of GIS seem well suited to provide the needed approach to integrate environmental, natural resource, cultural, and economic data in order to analyze possible impacts.

## Geographic information systems: their use and constraints

Baseline studies of the natural resources within a project area, including land use surveys, soil surveys, vegetation studies, and forestry studies, are typical requirements of an environmental assessment of an economic development project area. Primary data on these resources can be acquired by combining various map and imagery sources, if available, through GIS functions. GIS provides its user with the means to integrate data and examine interrelations between natural resources and human activities. GIS also provides the ability to develop 'what if' scenarios. Experimentation with various development scenarios can help determine the impact of a development project on the existing condition. These GIS functionalities appear well suited to aid the environmental assessment process.

GIS has been used successfully, and increasingly, by individuals and agencies in wide ranging disciplines to provide information in support of various environmental policy decisions. With success, GIS and the analytical capabilities it provides finds its way into the work places of the non-geographer or mapping scientist. These users may have little understanding of the intricacies of analyses GIS makes possible, nor how the inherent limitations of source data used in these analyses will effect the reliability of the resulting information.

Unwary use of GIS, and particularly the unwary use of unreliable information resulting from analysis of poor quality data, is a significant concern. These concerns increase when the consequences of decisions based on GIS information are so potentially costly, environmentally, socially and economically.

Considerable research has been done on the spatial data quality and GIS error propagation issues referred to in this report. Results of these studies seek to raise awareness to the issues, and commonly, and appropriately call for additional research to develop error models (Veregin, 1990) and methods for assessing data quality (Mead, 1982). Research at NCGIA/Maine on the visualization of spatial data quality (Clapham and Beard, 1991), and the 'fitness for use' studies at the University of Washington (Chrisman and Lester, 1991) are current examples of these efforts. Nevertheless, the statement made by Goodchild in 1990 still holds, "the GIS systems currently available have basically ignored the fact that most spatial data used in these precise computerized analyses are not always perfectly accurate". Nor will they ever be. However, today's GIS systems allow users to stretch data beyond its original intent (Goodchild 1991), and violate fundamental mapping conventions with no consideration of the limitations inherent in most source data used in the processes.

Continual technological advancements and constant market force necessitate studies that reach beyond data quality and error propagation issues. Indisputably, the effects of inadequate spatial data, and the error introduced by computerization of these data are not to be ignored. However, research examining the effects of these limitations on the reliability and useability of information derived from the GIS analysis of these data, and its use in the decision process is noticeably lacking.

## Difficulties associated with GIS analysis

The fundamental limitations to GIS analysis are essentially fourfold. They are; the operational limitations which result from the computerization and analysis of geographic information, the inherent limitations in the source material used in analysis (Vitek et.al. 1984), the possibility that the user is unaware of these limitations, and the inability of current GIS software to inform users and if appropriate suppress potential misuses.

### Inherent Limitations

Cartographic generalization on maps used as GIS source data represent one limitation to GIS analyses. Maps produced at all scales are abstractions of reality. When producing maps at small scales cartographers are restricted in the amount of detail that can be portrayed. As a result of scale limitations, the cartographer begins to systematically (and sometimes unsystematically; referred to as cartographic license) omit detail from the source document when compiling the small scale map document. Enlarging the scale of the small scale map only produces an enlarged portrayal of generalized data. Detail and information omitted during the generalization process is not, and can not be recreated by simple enlargement.

The scale of the map used as GIS source should be consistent with the scale of the decision being made. For example, it is apparently inappropriate to use a 1:1,000,000 generalized vegetation map to determine vegetation cover at the town level. This warning may be elementary, but it was done in a recent World Bank pilot study because no other vegetation cover data was available. The point here is not whether this operation can be done with GIS, it can be. But, that it should be done with understanding of its affect on the resulting information. In the example mentioned, the vegetation data was used to demonstrate the capabilities of GIS, to show the need for larger scale data, and to promote discussion in the project planning process.

Other limitations in source material result from positional and attribute error, map projection distortions, the 'fuzzy boundary' problems associated with mapping physical features (e.g. soils, vegetation, geology), and the inability to determine the distribution of data within a thematic area.

**Operational Constraints**

The inherent limitations found in maps used as source material for GIS databases and analysis can be increased through data manipulation (Vitek, et.al., 1984). Operational limitations include errors introduced during digitizing, the numeric errors resulting from performing mathematical computations performed on single precision computers, the problems that arise when representing vector and polygon data in a raster format due to grid resolution constraints (Burrough, 1986), and the problems associated with polygon overlay, namely the creation, and miscoding of 'spurious' polygons that can result (Goodchild, 1978).

One of the major strengths, and commonly used purposes of GIS is to combine digitized map data from various sources, regardless of scale. Ironically, this function may also be one of its most problematic. The map overlay process allows users to integrate various types of spatial data into one data file and then perform queries on, or portray results from the newly created combined data. While the use of map overlay enables more extensive and more efficient analysis, the user must be aware of the problems associated with it.

Regarding the digital map overlay process Abler (in Goodchild, 1987) states:

*"Two GIS capabilities that excite enthusiasm among potential users are the ability to change map scales and the ability to overlay maps at random. Both capabilities are indeed exceedingly useful: they constitute much of the comparative advantage GIS holds over spatial analysis based on analogue maps. Both capabilities may also mislead decision makers who are unaware of the imprecision inherent in all cartography and who are untutored in the ways errors compound when map scales are changed or when maps are merged".*

To strengthen Abler's assertion, Dahlberg (1986) warns that integration and relation of map data from various unrelated sources using GIS is a uncertain proposition unless the data base has been assembled from sources of high quality. Unfortunately, most developing countries fall into the 'data poor' category.

Functionally, map "overlay uses position information (geometry) to construct new zones which share parentage from the separate layers" (Chrisman, 1990). The primary concern with overlaying polygon 'coverages' is that false sliver polygons with erroneous attributes are unintentionally created. These slivers result primarily from the generalized boundaries portrayed on maps, and also from variations in individual used to digitize maps used as source material.

Because of these inherent data and operational limitations, the user of GIS for environmental assessment in 'data-poor' regions faces a dilemma. While GIS provides a tool to effectively derive information, the reliability of decision ready information is directly related to the original quality of the data.

Acquisition of quality primary data (i.e., from field surveys tied to accurate survey measurements (GPS), or differentially rectified aerial photography) require time and money. These resources are at best scarce when developing countries are paying the bill and waiting for the project to begin. These constraints to GIS analysis can be controlled however, by acknowledging and understanding data and computerization limitations, and carefully planning data acquisition strategies. Unfortunately, it is more common that data quality issues are ignored or unconsciously overlooked in the project planning and data acquisition stage; due to time constraints, or lack of knowledge.

The issue here is not whether errors or inaccuracies in spatial data exist; they do. Rather the issue is to make non-technical GIS users aware of these limitations. Analysis and use of poor quality data suggests uncertainty in the results of GIS. This uncertainty increases the risk associated with decisions made from these results. GIS has been used effectively and successfully in applications where source data is available and of traceable lineage and quality. There must be however, concern regarding the reliability and even the validity of GIS analysis performed on data of poor quality.

**Information Reliability and Decisions**

To further complicate the situation, and to suggest a more significant limitation of today's GIS systems — they do not provide the means for determining the reliability of information they produce. Unfortunately, the computerization of spatial analysis with its rapid response may imply, to the unaware, an unfounded belief of unquestioned reliability in the output of these systems.

Because of these characteristics, a real problem occurs when trying to determine the ultimate value of GIS information in the decision process. GIS/Geography/mapping science professionals have been unable to convince the typical GIS users as to the importance of data quality and operational limitations, and how these factors can constrain the GIS process. There are at least three reasons for this inability. We are unable to keep up with the rapid increase in GIS users from non-mapping backgrounds, the issues of data quality and information reliability are seen by many users as a negative response to the overwhelmingly positive accounts of GIS application, and despite the quality and abundance of papers dealing with these issues, we have been unable to express their significance in terminology that is familiar to users with different disciplinary backgrounds.

## Discussion

GIS can be used to collect, integrate, and analyze diverse spatial data to provide information on a variety of development issues; whether they be evaluating the regional impact on natural resources of economic development projects (Jagannathan, et.al., 1990), assisting land use mapping in the Philippines (Luscombe, 1990), supporting coastal resource planning and management in the Mediterranean (World Bank/European Investment Bank, 1990), facilitating land titling in Thailand (Williamson, 1986), or providing information to assist environmental assessments of development projects. However, a viable solution to decreasing the environmental impacts of development projects will only result from meaningful, responsible, and defensible policy, which GIS can merely influence. What GIS provides is a new view of existing geographic data, including imperfections. It is the GIS analyst's responsibility to understand the limitations of the information resulting from GIS, and to apprise the decision-maker as to its reliability so that it is used in an appropriate, responsible and defensible manner.

While more complete and reliable information seems to be important, another question arises. Does better' information really provide for better natural resource, environmental, land use and development policy decisions? Reflecting on what is still the ultimate question, Zwart (1986) refutes the contention that better information will automatically result in better environmental decisions due to the complexity and multifaceted nature of these issues. In response, Niemann (1987) called for broad institutional change, and efforts towards technology transfer and departmental cooperation in data sharing as the key to better use of the information which GIS can provide.

These views come from the developed world. In the context of developing countries, the words multifaceted and complex do not begin to define the difficulties confronting economic and environmental policy making. And the social, economic, and political obstacles to institutional change, and departmental cooperation may be beyond definition. It is in these conditions that the use and value of GIS will encounter a staunch test.

The real concern in developing countries is upgrading human life. The authors question whether GIS can have an immediate and direct affect on these concerns without extreme institutional changes. However, we feel if used appropriately, GIS can provide far more information on project areas than previously available, as well as information regarding the potential impact of development projects. The issue is not simply the use of GIS for technology sake, but the use and value of the information provided by GIS technology in making decisions. We believe that if used with an understanding of its limitations and if the 'right' questions are asked, GIS can provide information to illuminate illogical and indefensible decisions. And perhaps, the information that results will provoke the necessary institutional changes.

The task ahead for GIS advocates is to continue fostering interest in GIS. We should continue demonstrating the potential uses of GIS and related technology to a variety of disciplines. Concurrently we must educate ourselves, and prospective GIS users to the existence of limitations in source material, as well as the operational constraints of GIS. If we are successful in understanding and fostering awareness to these issues in the ultimate user, the decision-maker, then together we can discuss and determine conditions for information use. The decision-makers themselves may have the ultimate knowledge as to the required level of information reliability. They are presumably more aware of the gravity of the decision being made, the degree of defensibility and responsibly required, and the amount of risk they are willing to assume.

These are not easy tasks. To be successful, it will require continued effort on the part of GIS professionals to cross disciplinary lines as advocates of the technology. It will require GIS professionals to understand the constraints and to explain these constraints in an understandable manner to a variety of scientists and policy makers. Ultimately the success of GIS will depend on the success of the decisions themselves.

## Conclusion

Dr. David Simmonett suggested several years ago that development decisions in developing countries are made in a atmosphere of great uncertainty due to the weaker information base found there (in Morgan, 1990). As a consequence of this data poor condition, sometimes undesirable economic, cultural and environmental decisions occur. Yet, regardless of this lack of information, development projects continue to be identified, prepared, and implemented (Luscombe, 1986).

The much heralded advancements in GIS technology and its influence in changing the world has yet to do much to change the situation in developing countries. As tersely stated by a GIS software distributor from Sri Lanka, "developed countries are using GIS to organize what is already organized, it will prove more difficult to use GIS where things have yet to be organized".

Using GIS for analysis and development project impact modeling can be an effective instrument for providing information to aid the environmental assessment process. While the extent of technology used is ultimately dependent on the specific project, the utility of GIS is undeniable. However, due to the 'data poor' conditions commonly found in developing countries, the results and reliability of GIS analysis may be limited. Constraints arising from this condition make results of these analyses less valuable and possibly misleading if used as a definitive planning tool. Therefore, the authors believe the results of GIS analysis of 'data-poor' regions should be used primarily as a means to raise awareness, encourage discussion, and facilitate expert interaction in the project planning and environmental assessment process of development projects.

## References

Brooner, William G., Paul Maughan and David Thibault. 1988. "An assessment of information demands for remote sensing and geographic information system technologies". Final Report Submitted to the United States Agency for International Development. Earth Satellite Corporation. Chevy Chase, Md.

Bromley, Daniel, 1986. *Natural Resource Economics; Policy, Problems and Contemporary Analysis.* Hingham: Kluwer-Nijhoff.

Burrough, Peter A. 1986. *Principles of Geographical Information Systems for Land Resource Assessment.* Clarendon Press. Oxford.

Changchui, He, Ryerson, Robert, and Barry Haack. 1991. "The UNDP/ESCAP Regional Remote Sensing Programme for Asia and the Pacific". *Technical Papers 1991 ACSM/ASPRS Annual Convention,* vol.3.

Chrisman, N.R., and Marcus Lester. 1991. "A Diagnostic Test for Error in Categorical Maps". *Technical Papers 1991 ACSM/ASPRS Annual Convention. Auto-Carto 19.* vol.6.

Chrisman, N.R. 1990. "Modeling Error in Overlaid Categorical Maps". *The Accuracy of Spatial Databases.* ed. Michael Goodchild and Sucharita Gopal. Philadelphia, PA. Taylor and Francis.

Clapham, S.B. and Kate Beard. 1991. "The Development of an Initial Framework for the Visualization of Spatial Data Quality". *Technical Papers 1991 ACSM/ASPRS Annual Convention.* vol.2.

Dahlberg, Richard E. 1986. "Combining Data from Different Sources". *Surveying and Mapping.* 46: 141-149.

Dale, Peter F. and John McLaughlin. 1988. *Land Information Management: An introduction with special reference to cadastral problems in Third World countries.* Clarendon Press. Oxford.

Fisher, Peter. F. 1990. "Knowledge-based Approaches to Determining and Correcting Areas of Unreliability in Geographic Databases". *The Accuracy of Spatial Databases.* ed. Michael Goodchild and Scharita Gopal. Philadelphia, PA. Taylor and Francis.

Goodchild, Michael F. 1991. "Keynote Address, Symposium on Spatial Data Base Accuracy". *Proceedings, Symposium on Spatial Database Accuracy.* Melbourne, Australia.

Goodchild, Michael F. 1990. Preface of, *The Accuracy of Spatial Data Bases.* ed. Michael Goodchild and Sucharita Gopal. Philadelphia, PA. Taylor and Francis.

Goodchild, Michael F. 1989. Accuracy of spatial databases: initiative 1 specialist meeting report. Technical Paper 89-1. National Center for Geographic Information and Analysis. Santa Barbara, CA.

Goodchild, Michael F. 1980. "The Effects of Generalization in Geographical Data Encoding". *Map Data Processing.* ed. Herbert Freeman and Goffredo G. Pieroni, 191-205. New York. Academic Press, Inc.

Goodchild, Michael F. 1978. " Statistical Aspects of the Polygon Overlay Problem". In *Harvard Papers on Geographic Information Systems.* ed. G. Dutton. vol.6. Addison-Wesley, Reading, Mass.

Hassan, Hassan., Luscombe, B., Wayne, and Villagran, Felipe. 1991. "Strategies for Coordinating Natural Resource Information Management Programs". *Technical Papers 1991 ACSM/ASPRS Annual Convention.* vol.3.

Jagannathan, Vijay, Hideki Mori and Hassan Hassan. 1990. "Research Proposal. Use of Geographic Information in Natural Resource Management". Internal World Bank Office Memorandum. The World Bank, Washington, D.C.

Jagannathan, Vijay, Hideki Mori and Hassan Hassan. 1990. "Applications of Geographical Information Systems in Economic Analysis; A Case Study of Uganda" *World Bank Environment Department Working Paper.* The World Bank, Washington, D.C.

Luscombe, B. Wayne. 1990."Land Use Mapping in the Philippines using SPOT Satellite Imagery". *Satellite Remote Sensing for Agricultural Projects.* World Bank Technical Paper Number 128. ed. J.P. Gastellu-Etchegorry. The World Bank, Washington, D.C.

Luscombe, B. Wayne. 1986. "Spatial Data Handling in Data-Poor Environments". unpublished PhD. Dissertation. Simon Fraser University.

Mead, Douglas A. 1984. "Assessing Data Quality in Geographic Information Systems". *Remote Sensing for Resource Management.* ed. C.J. Johansen and J.L. Sanders. Soil Conservation Society of America. Ankeny, Iowa.

Morgan, Glenn. 1990. "Remote Sensing Activities in the World Bank: a Review of Experiences and Current Technical Capabilities". *Satellite Remote Sensing for Agricultural Projects.* World Bank Technical Paper Number 128, ed. J.P. Gastellu-Etchegorry. The World Bank, Washington, D.C.

Morgan, Glenn. 1989. "Geographic Information Systems in Development Planning: a World Bank Perspective". Paper prepared for GIS '89. Vancouver, Canada.

Niemann, Bernard J. 1987. "Better Information for Better Decisions: No Question About It". *Proceedings of URPIS 15*. Australian Urban and Regional Information Systems Association. Hobart, Tasmania.

Paul, Charles K. 1989. "Mapping in the U.S. Foreign Assistance Program". Paper presented at the *American Society of Civil Engineering Specialty Conference on Surveying Engineering: Managing New Technologies*. Denver, Colorado.

Veregin, Howard. 1990. "Error Modeling for the Map Overlay Operation". *The Accuracy of Spatial Data Bases*. ed. Michael Goodchild and Sucharita Gopal. Philadelphia, PA. Taylor and Francis.

Vitek, J.D., Stephen J. Walsh and Mark S. Gregory. 1984. "Accuracy in Geographic Information Systems: An Assessment of Inherent and Operational Error". *PECORA IX Symposium*.

Walsh, Stephen J., Dale Lightfoot and David Butler. 1987. "Recognition and Assessment of Error in Geographic Information Systems". *Photogrammetric Engineering and Remote Sensing*. 53 (October).

Williamson, Ian. 1986. "Cadastral and Land Information Systems in Developing Countries". *The Australian Surveyor*. 33 (March): 27-43.

World Bank. 1990. *The Environmental Program for the Mediterranean; Preserving a Shared Heritage and Managing a Common Resource*. The World Bank / European Investment Bank.

World Bank Operational Directive 4.00, Annex A: Environmental Assessment. August 1989. The World Bank, Washington, D.C.

Zwart, Peter. 1986. "Better Information for Better Decisions: Fact, Fiction or Maybe". *Proceedings of URPIS 14*. Australian Conference on Urban and Regional Planning Information Systems. Melbourne.

[1] *The term "reliable information" is used in this report to describe the result, or product of analyzing quality data with GIS. The reliability of information is a direct result of the quality of data used in analysis, and has a direct impact on how the information can, and will be used in the decision process.*

[2] *The term "data poor" is used to describe a general "data environment" condition where a lack of quality data exists.*

[3] *Realizing the word "better" is subjective, the authors use it to mean improved, logical and equitable. It should be noted that the authors believe one of the primary motivations for using GIS is to help reveal illogical and/or self-justifying (primarily land use) decisions.*

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Using GIS and image processing to prioritize cumulative effects assessment[1]

**Stephen Bernath** (presenting author)

Field Operations Coordinator

**Matthew Brunengo**

Environmental Geologist

**Stuart Smith**

GIS Analyst
Forest Practices Division
Washington State Department of Natural Resources
P. O. Box 47012
Olympia, WA 98504-7012, USA
Tel: (206)753-5315
Fax: (206)586-7311

**Lisa Lackey**

Remote Sensing Analyst
Pacific Meridian Resources
5200 S.W. Macadam Avenue
Suite 570
Portland, OR 97201, USA
Tel: (503)228-8708
Fax: (503)228-8751

## Abstract

*Concern for the potential cumulative effects of forest practices on wildlife, hydrology, fish, and erosion/sedimentation has become critical throughout the West. In an effort to analyze the cumulative effects of forest practices in Washington, the state's Department of Natural Resources has initiated a project to assess the potential for cumulative impacts for each forested basin across the state. Because detailed assessments in basins or across the landscape cannot occur simultaneously state-wide, a system of priorities for identifying basins that are most susceptible to cumulative effects has been developed using GIS and image processing technologies.*

*For example, harvest in the rain-on-snow zone of the Pacific Northwest can change the runoff characteristics of a basin affecting channel morphology, water quality, and fish habitat. The water available for runoff can be calculated from (1) NOAA 10-year 24 hour precipitation data, (2) rain-on-snow information, and (3) classification of the size and crown closure of forest stands based on interpretation of Landsat imagery. Such data can be used to prioritize basins as to the risk of sustaining adverse cumulative effects. Management and/or regulatory strategies for harvest can be formulated to avoid significant environmental impacts.*

## Background

Definition — Cumulative effects (in the forest) are the changes to the environment caused by the spatial and temporal interaction of natural ecosystem processes with the effects of two or more forest practices.

Washington's State Forest Practices Board[2] (FPB) has recognized concerns regarding cumulative effects (CE) for more than 10 years. In 1982, in an effort to begin to address CE, Ecosystems, Inc. was hired by the FPB to prepare a document that presented current knowledge of CE with respect to forest practices (Geppert et al., 1984). A second task in the original request for proposal was to conduct basin examinations in Washington relative to CE; this activity never began.

In the following three years pressure mounted for further regulation of forest practices. The pressure manifested itself in legislative power struggles between the interested parties, wars between experts in front of the FPB, and litigation. In an effort to curtail these divisive activities, the Timber-Fish-Wildlife (TFW) agreement (1987) was negotiated by four interest groups: Indian tribes, state agencies, forest landowners, and environmental organizations. This agreement ultimately changed the Forest Practices Act[3], the forest practices regulations[4], and inspired cooperative efforts, including joint research and a commitment to work together on the ground. However, the agreement and subsequent changes in the regulation of forest practices did not result in a process or method for dealing with CE.

In 1989, the TFW cooperators approved an issue paper (Golde et al., 1989) recognizing that CE should be addressed comprehensively. That effort resulted in asking TFW's Cooperative Monitoring, Evaluation and Research (CMER) committee to focus more of its energy into research that would fill the holes in our knowledge regarding CE. In addition, the FPB had not addressed CE specifically in the regulations. This limited the conditioning of forest practice applications for CE (Olson, 1989) by the Department of Natural Resources[5] (DNR).

Two events occurred in 1989 prompting action in addressing issues that had not been resolved in TFW. The first was a lawsuit concerning CE as a result of proposed harvest around Lake Roesiger. The other was a second round of negotiations, known as the Sustainable Forestry Roundtable (SFR).

During the SFR negotiations, the same parties that participated in the TFW Agreement plus a new interest group, local governments, asked CMER: "Given current knowledge, what kind of technical approach could be used to begin dealing with cumulative effects?" CMER responded to the request with an initial report to SFR regarding how CE might be addressed (1990). While SFR incorporated this information into their proposal (1990), the negotiations ultimately failed. Several attempts were made during the 1991 legislative session to codify some of the concepts contained in the SFR proposal, including analysis of CE, but these, too, were unsuccessful. What the Legislature did provide was incentive. If the FPB would, by July 1, 1992, adopt regulations that address CE, wetlands, wildlife, rate of harvest, and clearcut size and timing, then dollars for implementation of the new rules would be made available.

To identify a methodology for approaching CE, the DNR, in consultation with CMER, was directed to complete the work originally initiated by SFR and to provide a scientifically-based tool for the FPB by March 1, 1992 (WAC 222-16-046). This effort has culminated in a proposed framework for dealing technically with CE on a watershed basis for state-regulated forest lands in Washington (12 million acres).

## Cumulative effects assessment[6]

Currently, regulatory decisions on forest practices rely mostly on professional knowledge and evaluations regarding *local* conditions, i.e., the state of the forest-practice unit and areas directly adjacent to it. The ecosystem-and basin-wide perspectives are typically absent. Similarly, the resource information necessary for assessing existing or potential cumulative problems, the methods of analyzing the information, and standardized means of communicating the results among interested parties are generally inadequate or nonexistent.

A basic premise behind CE assessment is that landscapes differ in their sensitivities to forestry activities. Hydrologic and geomorphic processes vary spatially with differences in climate, geologic materials, structure, and terrain; vegetation and habitat conditions are partly controlled by these processes and by the history of natural and human-caused events in the forest. Furthermore, the broad class of activities encompassed in 'forest practices' has an equally broad spectrum of potential effects on the environment.

Because of the diversity of operations and conditions, all areas are not equally sensitive to any particular forest practice, and probably no area is sensitive to all possible negative effects. Consequently, the risks to water quality, fish and wildlife habitat, and public improvements associated with the range of management activities vary from place to place.

The essence of our approach is to develop a more explicit and objective procedure for performing and displaying basin-scale environmental evaluations. The results will be used by managers to design local strategies to address identified CE problems.

## The design

The CE assessment examines basins for existing ecosystem conditions, actual and/or potential vegetation changes, accelerated erosion, increased flooding, deterioration of water quality, and changes in stream morphology. Many of these factors are evaluated with respect to fish and wildlife habitat needs. Resource impacts resulting from both natural background conditions and past human activities are considered.

The assessment of basin-wide hazards and the forestry activities likely to affect them can assist managers in minimizing both local and cumulative problems by aiding in the design of watershed-specific harvest plans. Subsequent operations may incorporate best management practices specified in the forest practices regulations, or alternative measures where appropriate.

Evaluation of risk to public resources helps in establishing appropriate decision criteria based on water- and habitat-quality standards. Appropriate conditioning of forest practices to minimize the potential for significant adverse effects is the result. In addition, the assessment can be used to identify opportunities for monitoring to determine the effectiveness of management and regulatory actions in meeting public-resource objectives.

It is intended that CE assessment will eventually take place in all basins containing state and private forest land. However, staff and funds are not available to perform all work on all basins in a short time. Therefore, the CE assessment has been organized into two phases of analysis. Level 1 analysis is intended to be performed over approximately six months. Existing information, regional staff, and local expertise will be used to look at all the basins of the state containing forest lands. Level 2 analysis will be performed using 'certified' watershed analysis teams, funded by the legislature or provided by others (such as landowners or tribes). This analysis will deal more intensively with basins of concern, gather more information, and identify critical decision criteria.

Watershed screening will determine the order in which the basins will be examined, regionally for Level 1 analysis and state-wide for Level 2 analysis, by prioritizing them on the basis of broad-scale environmental factors. The prioritization will not preclude others from performing Level 2 watershed analysis in watersheds of concern when desired. Neither will the initial prioritization prevent potential adjustments to the state-wide priorities as Level 1 is performed. Watershed screening will provide some information, not previously available, to the CE assessment in each basin.

## Designation of basins

The first step was to delineate basins for the purpose of screening and analysis. It was presupposed that the basins should be contained within the water resource inventory areas (WRIAs[7]), and be fourth- to fifth-order in size. Adjustments were made to the latter criterion based on physiography, land use, and ownership. Using these criteria, 204 basins were designated, with one to seven basins in each WRIA. (Designation of these basins does not preclude further subdivision for detailed analysis.) Approximately 30 of the basins will be excluded from screening because they do not contain significant tracts of commercial forest. The boundaries of these basins have been digitized and are reproducible from DNR's GIS.

## Prioritization of basins

The remainder of this paper is focused on how GIS and remote sensing are being used to prioritize basins for CE assessment across the state. To prioritize basins on a regional and state-wide basis in a timely fashion, information had to be easily obtained from maps, GIS, or satellite imagery. The following screens[8] were identified for key categories of hazard and risk.

### Slope instability screen

The assumptions underlying the slope-hazard screen included that it be based on physical factors relevant to slope stability, that it use available information, and that the information is in forms that can be incorporated into a general instability rating.

Existing slope-hazard ratings are available for state and private commercial forest lands in the soils layer stored on the DNR GIS; the ratings for national forest lands are being obtained from the US Forest Service. Information on slope gradients is being obtained by purchasing sets of digital elevation data from the US Geological Survey and processing them to produce slope maps.

A storm-input layer is being developed to reflect the role that storm precipitation plays in triggering landslides. The isohyetal map of 10-year 24-hour precipitation (from NOAA) has been digitized, and a precipitation zones map has been developed. These will be combined into a model rating of potential storm-water input (see Hydrologic Screen section below).

The combination of the slope-hazard ratings and slope gradient information will show where steep and unstable areas are located. The storm-input layer will overlay the soils-slope layer to show where storms could be most effective in causing slides. The resultant combinations will be given numerical slope-hazard ratings.

To validate the ratings and add information regarding potentially unstable rock types, landforms, and known problem areas (information not otherwise contained in the slope-hazard ratings), teams of geologists and soil scientists familiar with different areas of the state will be brought together. These experts will use aerial photography, geologic maps, reports on historic slides, and personal experience to check the designations. This will provide verification that the identified 'unstable' zones are appropriate, and allow for modification of ratings with additional information.

### Fisheries screen

The two criteria recommended for the state-wide fish screen recognize the importance of (1) streams that supply water to fish hatcheries, and (2) streams that support the state's natural/wild fish populations.

The value of including waters supplying fish cultural facilities is self-evident, since production is fully dependent on reliable, high-quality water. The inclusion of the proportion of utilized habitat, indicated by the presence of fish species in the basin by river mile, will identify where the habitat is being used.

The Washington Rivers Information System (WARIS), a GIS data base maintained by the Washington Department of Wildlife, contains site-specific information on state, tribal, and federal hatchery facilities and the presence and absence of fish species. These data have been further updated and verified, and currently resides on the DNR GIS.

### Wildlife screen

The wildlife screen will score each WRIA for five factors: (1) the proportion of the forested area* in late-successional habitat, (2) the proportion in functional (60-acre patches) late-successional habitat, (3) the proportion in large patches (640 acres) of late-successional habitat, (4) the distribution of 60-acre patches, and (5) the proportion of vegetative cover in mid-successional forest.

The use of late-successional forest in the wildlife screen focuses on providing diversity of habitat for species not found in earlier successional stages. The criteria regarding size and distribution of patches are incorporated because these factors influence the dispersal of wildlife. The consideration of the mid-successional stage reflects the need for preserving natural diversity in a basin by identifying the habitat that is between early and late successional stages.

The classification of vegetation for wildlife habitat types is a result of processing Landsat images, as explained below. The evaluation of the classifications for wildlife will be performed using the ERDAS (Atlanta, Georgia) Clump and Sieve algorithm.

### Hydrologic screen

The primary hydrologic problem related to forest management in Washington is the impact of timber harvest on runoff during rain-on-snow events. In the Pacific Northwest, the heaviest rains generally occur when cyclonic-frontal winter storms bring warm, moist air from the southwest. Since there is likely to be some snow on the ground at middle and higher elevations when these storms occur, snowmelt can combine with rainfall to aggravate floods and trigger landslides. The distribution of forest stands with varying degrees of hydrologic maturity (i.e., their ability to function like mature forests with respect to snow accumulation and melt processes), and their similarity to stands with natural runoff characteristics, will be used to rate the basins for the potential to alter peak flows and channel morphology.

To perform this exercise, several GIS layers are being adapted or developed for an analysis that is based on modeling of hypothetical rain-on-snow storms. The 10-year 24-hour storm isohyets were digitized from the NOAA atlas (Miller et al., 1973). Using data from the National Weather Service and cooperative snow surveys, a map of five precipitation zones was created to delineate areas likely to have various amounts of snow on the ground (in early January). The peak rain-on-snow zone was defined as the area in which the amount of snow available is approximately the same as could be melted by the 10-year 24-hour rainstorm (assuming reasonable temperature and winds). Below this peak rain-on-snow zone are the lowlands (little snow on the ground) and rain-dominated (some snow, but rain is the primary source of water input) zones; above are the highlands (too cold, so little liquid water generated) and snow-dominated (some snowmelt, but delay during percolation through the snowpack, zones. Although the distribution of these zones is controlled primarily by elevation, other climatic and geographic factors (e.g., latitude, aspect, storm tracks, mountain ranges) also affect their locations. These were reflected in the mapping.

The hydrologic maturity of forest vegetation is being identified through processing of Landsat imagery (as identified below). Recovery of a young forest to hydrologic maturity occurs when the properties controlling snow accumulation and melt, such as canopy closure, match those of a mature forest. These characteristics, adjusted for region, species, etc., can be used as standards of maturity.

Using these data layers, an approximation of the amount of water available for runoff from a hypothetical rain-on-snow storm can be generated using ERDAS GISMO. The 10-year 24-hour precipitation amounts are combined with the amounts of snowmelt appropriate for each zone (none in the highlands and lowlands, a maximum amount in the peak R/S zone) and each vegetation type (more snow available and faster melting in hydrologically immature stands and nonforested areas). The GIS calculates the areally-weighted averages of storm precipitation plus snowmelt for each basin under two sets of assumptions: (1) all forest lands support hydrologically mature vegetation (i.e., fully-forested conditions); and (2) the pattern of vegetation is as interpreted from the Landsat imagery (i.e., current conditions). The difference between these calculated values for a given basin represents the change in water available for runoff that could occur in a storm like the one modeled, as a result of timber harvest: the greater the difference (say, over 2-3 in. of increased water over a whole basin), the greater the potential for increased peak flows and damaging effects downstream.

Thus, the hydrologic screen will score each basin by calculating the change in water available for runoff from a mature forested condition to current conditions, for a model 24-hour rain-on-snow storm.

### Vegetation classification

The vegetation classifications for both the wildlife and hydrologic screens are being performed under contract by Pacific Meridian Resources. The classes include: (1) late-successional conifer stands, (2) mid-successional conifer stands, (3) early-successional conifer stands, (4) other forest lands, and (5) water. These classifications have been processed through the use of geocoded and terrain-corrected 1988 Landsat TM imagery (EOSAT, Lanham, Maryland) and ERDAS image-processing software.

Six national forests and three national parks have already been mapped using more detailed vegetation classification of 16 and 31 categories, respectively (Teply and Green, 1991); these public lands have been recoded to fit the five-class scheme. The remaining forested areas are being classified using an unsupervised classification in combination with aerial photo interpretation, field reconnaissance, and the expertise of DNR regional foresters. The resulting classification is filtered and polygons less than five acres eliminated. The wildlife and hydrologic analyses are then carried out for each WRIA.

The wildlife screen will utilize the late-successional and mid-successional vegetation types of the forested areas of Washington. The hydrologic screen will consider mid- and late-successional vegetation classes to be hydrologically mature.

## Summary

Attempts at evaluating cumulative effects have been continuing in Washington for over ten years. In response to concerns before the FPB and litigation, DNR, in consultation with its TFW partners, has constructed a framework for performing CE assessment on state and private forest lands. Such assessments could take from five to ten years state-wide. In order to prioritize where cumulative effects assessment should take place first, a watershed screening process is being utilized. GIS and remote-sensing techniques are being employed to gather and analyze broad-scale environmental data to prioritize basins. This exercise serves as an example of how data that are already resident on a GIS system, or that are easily obtainable, can be used in a timely fashion to identify where a CE assessment might be focused.

## References

CMER Task Force (1990). Interim Report to the Sustainable Forestry Roundtable. 18 pp.

Cumulative Effects Assessment Process (1991). Report to the Forest Practices Board by DNR and TFW. 9 pp.

Geppert, R. R., Lorenz, C. W., and Larson, A. G. (1984) Cumulative Effects of Forest Practices on the Environment - A State of the Knowledge. 208 pp.

Golde, M., Mumper, D., Gustavson, R., Olson, A., Geppert, R., McDonald, D., Somers, D., and Ross, R. (1989). Status and Recommendations on Cumulative Effects of Forest Practices. 17 pp.

Miller, J.F., Frederick, R.H., and Tracey, R.J. (1973). Precipitation frequency atlas of the western United States. Volume IX - Washington. National Oceanic and Atmospheric Administration, Atlas 2. 43 pp.

Olson, A. (1989). DNR Guidelines for Conditioning Forest Practices Applications. 16 pp. Sustainable Forestry Roundtable (1990). Proposal. 65 pp.

Teply, J. and Green, K. (1991). Old Growth Forests: How Much Remains? GeoInfo Systems. Vol 1. No. 4. Aster Publishing Corporation, Eugene, Oregon. p. 22-31.

Timber/Fish/Wildlife Agreement (1987). A Better Future in Our Woods and Streams - Final Report. 57 pp.

Washington Administrative Code (WAC) 222-16-046 (1990). Watershed Analysis Implementation. 2 pp.

## Footnotes

[1] The opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of any participant in, or committee of, the Timber/Fish/Wildlife agreement, the Washington State Forest Practices Board, or the Washington State Department of Natural Resources; nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

[2] The Washington Forest Practices Board was created in 1974 under the Forest Practices Act to promulgate forest practices regulations.

[3] The Forest Practices Act was passed by the state legislature in 1974 establishing a policy of protecting the public resources of the state coincident with the maintenance of a viable forest products industry.

[4] Best management practices and rules for administering the Forest Practices Act.

[5] The state agency responsible for administering the Forest Practices Act.

---

[6] Parts of the information in the following sections were taken from a report to the Forest Practices Board by DNR and TFW (1991).

[7] Water resource inventory areas (WRIAs) are the hydrologic administrative units of Washington adopted by the Washington Department of Ecology.

[8] Screening, as used here, is the process of focusing where CE of a particular type may be of concern.

[9] The forested area in Washington has been determined through the use of USGS digitized land-use

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# The future of GIS in the regulatory process: The OSLO case study

**James W. Case**

CASE Biomanagement
#1142, 555 - 4th Ave. S.W.,
Calgary, Alberta, Canada T2P 3E7

**Judith A. Smith**

Concord Environmental
#1140, 555 - 4th Ave. S.W.,
Calgary, Alberta, Canada T2P 3E7.

## Abstract

*The regulatory process for approving proposed industrial developments, which involves the preparation of an Environmental Impact Assessment (EIA), can require the analysis of large amounts of spatial data.*

*A GIS was used in the preparation of the draft EIA of the proposed $3.4 billion OSLO project; an oil sand extraction upgrading facility in Alberta. GIS technology was considered vital to the analysis and presentation of the enormous amounts of data gathered for the EIA. It became apparent that in addition to improving the traditional manual EIA process, use of the GIS could improve data management, analysis and presentation, resulting in a better assessment and improved environmental protection. The GIS application could subsequently be used for on-going planning and management of environmental information.*

*Suggestions are also presented that the GIS could be used more extensively in the regulatory process.*

*A public "library" of digital environmental information could be built up over the years in much the way the ERCB well log data was accumulated. The availability of this information would reduce the cost of environmental studies and facilitate on-going monitoring and research by private and public agencies. The place of GIS in EIA and subsequent monitoring in developing countries is discussed.*

## Introduction

The OSLO Project is a proposed oil sands development in northeast Alberta. It will include a surface mine and extraction plant, which produces bitumen, located on a portion of OSLO Lease 31 near Kearl Lake, about 60 km north of Fort McMurray. In addition, an upgrader, which produces synthetic crude oil from the bitumen extracted from the oil sand, will be located in an area about 7 km from Redwater, within ready access of the city of Edmonton. The mine and extraction plant will be connected with the upgrader by pipelines, one to bring diluted bitumen to the upgrader, and another to return the diluent to the mine and extraction plant.

The project cost about $4.5 billion and is being undertaken as a joint venture by the owners of six oil sands leases in the Fort McMurray area; hence the name OSLO; an acronym of "Other Six Lease Operators". In terms of both geographic extent and database size, OSLO is one of the largest natural resource projects in the private sector.

As part of the regulatory requirements for the project, an environmental impact assessment (EIA) must be completed before approval can be given. An EIA is a study done to determine the effects a proposed project might have on the environment. It is done to ensure that the proponent of the proposed project will satisfy and evaluate all potentially significant environmental effects of the development. In addition, the proponent is required to give due consideration to means of avoiding or mitigating any adverse potential environmental effects.

In North America, the EIA process usually involves an approving agency listening to three concerned parties: the proponents of the project, the general public and the special interest groups. The proponent must prepare an EIA outlining all aspects of the project, the potential for impact on the environment, and the measures, if any, that will be taken to mitigate effects. It is common practice in Canada, U.S.A. and Mexico, for an EIA to be prepared for the proponent by a third party (i.e., consultants).

Over the past few years, some developers have been required to prepare environmental assessments that meet the requirements of more than one piece of legislation, and undergo joint reviews by more than one government body. This confusion arises because both the provincial and federal governments have responsibility for environmental resources in Canada. Had the OSLO project proceeded (it was postponed in June 1991), it would have undergone a joint provincial/federal hearing, because the project was provided federal financial support and would have effected resources that fall under federal jurisdiction (i.e., fisheries, migratory birds and navigable rivers).

## The EIA process

In Canada, the influence of environmental issues on project approval, and the process used to prepare EIAs are expected to increase in complexity over the next few years. There is also an increasing amount of confusion about which government body has the authority to review EIAs and authorize proponents of a project to proceed with its development.

The current EIA process can be unwieldy and impede development of some projects, while allowing other projects to proceed without a proper review of environmental effects. In addition, due to jurisdictional debates between governments, several developers have recently faced increased costs and time delays in project implementation.

The Canadian EIA process is also consultative in nature, and federal and provincial Ministers, acting in a discretionary capacity, are entitled to disregard the recommendations of appointed review boards (Ross 1991). In the past, behind-the-scene political decisions based on socio-economic issues have often taken precedence over environmental concerns in the project approval process. However, over the past 2 to 3 years, environmental concerns have more and more been treated as real issues, and the passage of the Alberta Environmental Protection and Enhancement Act (AEPEA) and the Canadian Environmental Assessment Act (CEAA) will legislate a more structured EIA review process that will encourage greater public involvement. Greater public participation will remove some of the weaknesses, such as political intervention, in the process, but this will mean that greater amounts of data will have to be processed and the presentation requirements will be more complex.

One of the major tasks in this process, with its many participants is managing the collection, correlation and analysis of the information generated by the EIA. This information must also be presented in a format that is meaningful and expressive to the participants and decision makers. There are many types of data required to perform an EIA. In most instances, field surveys are required to collect current information about wildlife, vegetation, fisheries, soil chemistry, terrain, historical resources, land use, etc. in the project study area. The extent of the field surveys depend on the availability of accurate, relevant, and contemporary information from existing sources.

Some of the obvious sources of information for an EIA are: municipal maps, government digital basemaps, air photo libraries, satellite imagery, and remote sensing. However, remote sensing information such as satellite surveillance data are usually under utilized. There are literally hundreds of other data sources available too. The difficulty rapidly becomes not one of too little information, but rather, which information to include in the analysis. For example, some databases may contain historically important but out-of-date information.

Two GIS technologies (Nucor on IBM PC and MapGrafix on Macintosh) were chosen to aid in the preparation of OSLO Project EIA. GIS technology, with its ability to manage, analyze and present spatial data in graphic and textual forms (i.e., maps and tables) has great potential as a tool in the EIA process. The analytical capabilities of a modern GIS, fully loaded with detailed information, make it feasible to make predictions and explore at least some of the consequences of particular kinds of projects on particular environmental issues.

The implementation of the GIS for the OSLO Project had four objectives:

- To assess the use of GIS technology to support of planning, design, construction, operation and decommissioning of the OSLO project.
- To bring all the environmental information into one manageable place.
- To facilitate objective assessment of impact hypotheses by means of data analysis.
- To produce illustrative figures for the impact assessment.

It was anticipated that the results of this exercise would provide a guide-line for the future use of a GIS in support of EIA.

In addition to allowing the rapid update of maps, the GIS would allow the complex analyses of spatial data sets that would be unfeasible using manual methods or CAD systems. It would be possible to combine many kinds of "maps", including existing paper maps, engineering drawings, architectural plans, digital maps, airphotos, and satellite images. It would also be possible to create new georeferenced information on the basis of data combinations and transformations.

The GIS-supported environmental EIA commenced in 1989 with extensive baseline inventories covering a region of nearly 1000 km². Environmental specialists conducted extensive field surveys to map existing resources (land use, vegetation, soils, terrain, wildlife habitat, historic resources, etc.) within a broad region surrounding the components of the proposed project. On the basis of their field surveys they produced detailed baseline maps. These baseline resource inventory maps were provided to us as paper maps, airphoto mosaics with polygons drawn on them, or as digital files from a variety of GIS and CAD systems.

Another component of the baseline database were the engineering drawings of proposed project facilities; for Lease 31 mine/extraction facilities and for the Red Water upgrader facilities. The engineering and mining groups within OSLO used AutoCAD and EAGLES to produce their facility plans.

Our endeavours quickly revealed that many people did not have a good understanding of what a GIS is and what its capabilities are. One of first tasks quickly became one of education so that realistic expectations were held by other team members. It was possible to taken advantage of the GIS analytical capabilities, however, because a specialist with extensive experience in both ecological disciplines and GIS was available to act as an "interpreter" between the various environmental specialists and the GIS technician. This was important as pointed out earlier, because the GIS is a new technology and while there may be a general understanding of GIS, the environmental specialists would have required considerable education, training and practice in order to use the system effectively on their own. The tight time constraints set by industry for the preparation of EIAs have traditionally not included budgets to educate the consulting team members in GIS concepts or to train them in the use of the particular GIS selected for use in the project. The procedure permitted them to get the analysis results they required within in the short time available and the format of the various products requested was consistent for all individuals requesting maps or analyses.

Production of analytical and map products was accomplished by following a procedure that evolved during the implementation of the GIS. The procedure involved carefully defining the end products, the data needed and the analysis procedures to be used to generate the product.

The end product required was defined jointly by the EIA specialist and the GIS specialist together. This insured that the proper product was generated and that the correct analyses were performed. The description included a rough sketch of the required map, the approximate extent, and major details, along with information to prepare a legend. The product description request also included the title, scale, name of the person requesting the map along with their address and phone number. Next, the data required was identified. If derived data were required, the source (factor) data sets were first identified. A step by step procedure is then developed with the technician, to produce the product using the various software modules and their associated functions.

The OSLO EIA GIS has now been archived so that it can be "taken out of the box and turned back on" when the OSLO Project is resurrected. At the time of the GIS archival, it contained 54 maps, 4 forestry data files, 1 habitat data file and the engineering drawings of all mining, extraction and upgrader facilities. The GIS archive contains files in the proprietary format of the GIS. In addition, the archive contains all the "raw" data import files so that the information can be transferred to another GIS if necessary.

Our experience with the use of GIS in the preparation of an EIA for a major project leads us to believe that GIS has great potential in enhancing the EIA process. However, this will require that GIS become an integral part of the EIA process rather than an "add on" to the existing traditional methodologies. The EIA process is going through an evolutionary change and now is the ideal time to develop the concept of GIS-based environmental management information systems.

In the remainder of this paper we will discuss the feasibility of using GIS technology to address the changing expectations for EIAs. It will address some of the ways we have used GIS in the environmental impact assessment of a very large project and ways it might be used in the future. The topics to be addressed are:

- Public involvement in the impact evaluation process.
- The scope and content of an EIA.
- Preparing a focused, comprehensive, technically-defensible and effective EIA.
- When should an EIA be required and who will review the EIA?

## Public involvement in the OSLO EIA process

OSLO initiated a new process for the assessment of the environmental impact. It was a scoping process and involved the following steps:

### Description

In this step, a clear understanding and description of the project developed. This was required to assess potential environmental issues.

### Issues Scoping

This step involved identifying the issues, assessing their impact and defining impact mitigation and monitoring.

### EIA Preparation

Preparation of the Impact Statements and the successful resolution of the issues identified were documented for review.

Key to the success of the issue scoping process was the involvement of key stakeholders (including federal and provincial government agencies, municipal governments and non-governmental organizations), as well as the public. In the case of the OSLO Project, substantial changes in the project description coupled with intensive stakeholder and public consultation resulted in a need to be able to quickly revise and regenerate maps and impact analyses based on new information or new wrinkles added to old information. It was believed that this could best be handled by the development of a GIS (Figure 1). We cannot stress enough the importance of establishing a procedure for routinely obtaining and incorporating updates to project plans and the environmental data base. This permits changes in the project engineering to be reflected quickly in the GIS. In what follows, the OSLO GIS is described in terms of its component parts and an overview is given of the various themes which have been used in the construction of the GIS.

Much of the information required for the EIA was developed as the result of issue scoping and hypothesis generation. The disturbance categories typically include, but are not limited to:

**Atmospheric disturbances** — disturbances to local and regional air quality.

**Ground disturbances** — disturbances to land resources caused by land clearing.

**Waterbody disturbance** — disturbance to water resources caused by diversions, flooding, siltation, etc.

**Public access** — increased access by the public to areas as a result of new roads, resulting in more traffic accidents, increased maintenance costs, increased pressure on sensitive resources, etc.

**Visual impact** — aesthetic degradation ("eye sores") resulting from project construction and operation.

The ability of the GIS to overlay several layers of georeferenced data makes it potentially valuable as a display tool for illustrating and analyzing impacts. The GIS can also be used to evaluate mitigation practices.

Once maps have been prepared for spatially oriented impacts, it is necessary to examine how each type of disturbances would affect the natural resource included in the database. Prior to the use of GIS, manual methods were used to compile these data on paper or film maps. These data were "coded" using symbols, patterns, and colours that were explained in a map legend or accompanying text. Survey data were presented in accompanying tables. Updating or expanding a survey required that all maps and tables be reconstructed from scratch. The ability to explore the effect of a different project configuration and produce a maps of the results was often cost prohibitive or, in some cases, technically impossible.

## Who should be involved in the impact evaluation process?

Based on our experience in the OSLO Project and other large projects, we can put forward the following comments regarding the advantage of using GIS technology in the EIA process. We have found that a GIS as a central component of the EIA process can [or could] lead to:

- the establishment of improved communication and understanding between the proponent team members, the regulatory agencies, and members of the public;
- the production of a more complete EIA, including a more complete, objective identification of all impacts and environmental safeguards, and the creation of innovative solutions to difficult problems. One of the most effective ways to be creative would be to assemble all the key stakeholders and use the GIS to devise solutions that will keep environmental impacts and economic feasibility of the proposed project within acceptable limits. The availability of an GIS at such meetings would permit the consequences of different scenarios to be examined. The requires that the GIS be designed and implemented at the start of the project and not, as has been the case in many projects in the past, concurrent with preparation of the EIA document. This will also help define the extent of the region to be studied and identify what and how to gather the data.
- The use of GIS could also result in the minimisation of project costs (intervenor costs) and time delays. By having maps and facility plans in a GIS, areas can be measured and maps updated much

more quickly than by manual methods. The potential also exists to improve the way in which the intervenors get involved. By making available all the environmental data in digital format, the redundancy of data and duplication of effort could be greatly reduced, as would the waste of funds. Each party could still conduct their own analysis but everyone would be starting with a more complete database than they could economically assemble by themselves. In addition, the use of GIS would require a change in the level of sophistication of the intervenors, which should result in better understanding of the project and higher quality interventions.

- If approved, the project would have in place, at the end of the process the basis for an environmental management information system which could help the company be diligent in enforcing their environmental code of practice by integrating the management of permits and licences, hazardous materials,

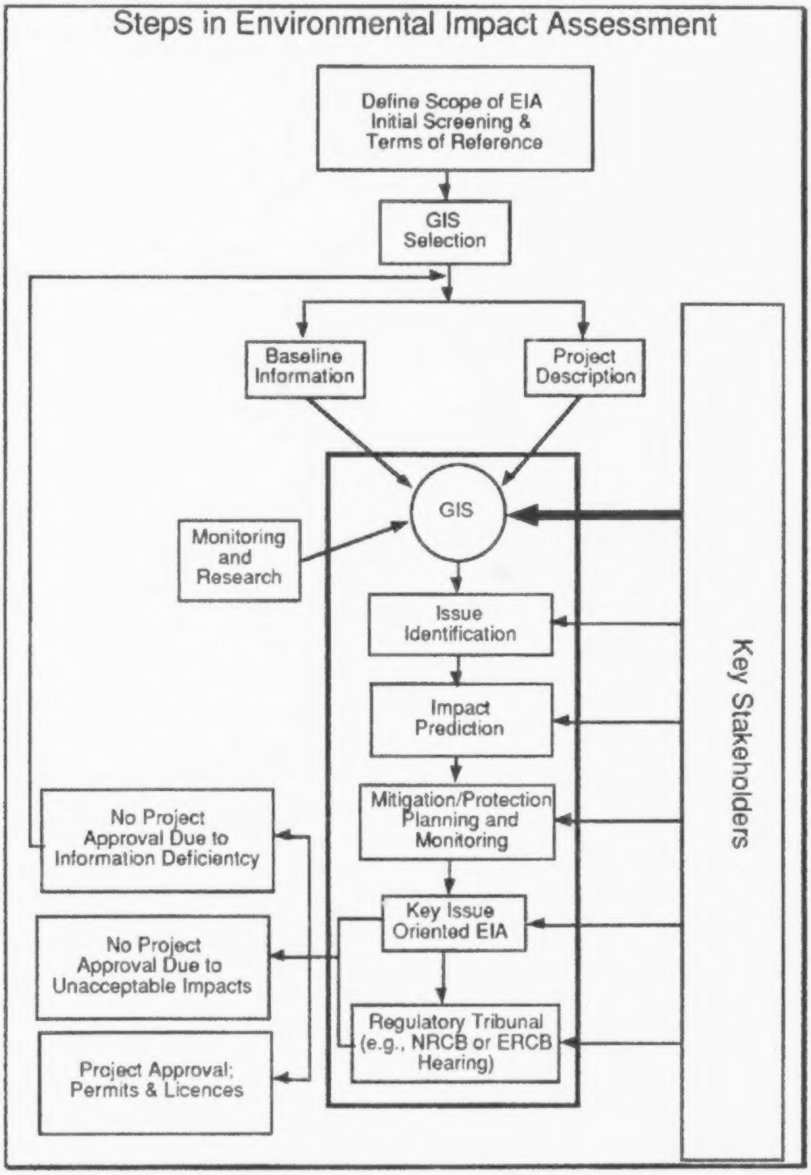

Steps in Environmental Impact Assessment
Define Scope of EIA Initial Screening & Terms of Reference
GIS Selection
Baseline Information
Project Description
GIS
Monitoring and Research
Issue Identification
Impact Prediction
Mitigation/Protection Planning and Monitoring
Key Issue Oriented EIA
Regulatory Tribunal (e.g., NRCB or ERCB Hearing)
No Project Approval Due to Information Deficientcy
No Project Approval Due to Unacceptable Impacts
Project Approval; Permits & Licences
Key Stakeholders

**Figure 1:** *The role of GIS in the Environmental Impact Assessment Process.*

toxic waste, environmental research, personnel training information, emergency response plans, compliance reporting, planning, etc.

- The assurance that environmental issues will receive appropriate review in relation to socio-economic issues, and that behind-the-scene political agreements do not determine whether a development is approved or disapproved.

A structured environmental management information system, with defined data manipulation procedures and reporting mechanisms, could also be audited. This presupposes the development of an "industry standard" defining the minimum requirements for data and recommended structure for the GIS. Each project would be subjected to a comprehensive systematic examination of its content to determine their individual environmental influence. Data submitted to public hearings needs to be stored in a library in a digital georeferenced format. Excellent models for such systems are the ERCB in Alberta which stores all well log data generated in Alberta or the Nature Conservancy GIS. The NRCB would seem to be an obvious candidate to develop and maintain such a "fee for use" environmental data library for Alberta; industry and government do not qualify as unbiased candidates. The digital information could be combined with existing data so that it would be possible to evaluate the influence of the proposed project on cumulative impact. The use of such an environmental information management system (EIMS) would have several positive effects:

- There would be an standard environmental management system which could be audited.
- Deficiencies would be much less likely and would be easier to detect and remedy. This would be especially useful in developing countries where experienced environmental impact assessment experts are not always available to conduct comprehensive and objective assessments.
- Data would be available in a more standardized format. This would make possible the concept of a public environmental data library.
- The development of a library of environmental data over the years would also address the problem of who pays for the off site environmental data collection if cumulative regional environmental impacts of a project are going to be required.
- Each project would be subjected to a comprehensive examination
- The EIMS provides the company with the basis with which to diligently carry out their environmental management strategy. It would also help them develop an environmental management strategy if they do not already have one.
- Each EIMS would augment the public environmental database without the shameful waste of mountains of paper; instead, the data would be truly "recyclable".

## The scope and content of an EIA

A frustrating aspect of the EIA process that each proponent faces is the confusion over how broad a scope and what issues the assessment should cover. In the past, only project components that were operated by one proponent were addressed in an EIA; other proponents were responsible for preparing separate and independent assessments. In the case of the OSLO project, the pipelines, power-lines, and municipal roads were not included in the EIA because they were being built by other companies or government departments. This is clearly unsatisfactory if we wish to integrate environmental information to monitor and plan for cumulative effects.

Recently, this fragmentary approach has been changing and, it is likely that in the future, all components of a project that are inextricably intertwined will be addressed in a single comprehensive document and review process. The proponents of the various components of such projects will either need to cooperate and work jointly to prepare one comprehensive EIA, or they will have to co-ordinate and conduct several EIAs so that they are completed within the same time frame using similar approaches. Standardized GIS procedures and design plans are seen as one means by which proponents could combine their efforts and resources to jointly accomplish this. The GIS would also facilitate analysis, systematic data collection, analysis, and presentation. This is particularly important because it would permit the cumulative effects of all components of a project to be evaluated.

Use of common or compatible GISs would allow proponents of all project components to submit their EIA information for incorporation into a "master" project EIA. This would permit the impact of the entire project to be assessed. In turn, making the project data available to a centralized data library would facilitate quantitative analyse of cumulative effects of the project in a timely manner by regulating agencies. All of this presupposes the rapid acquisition of GIS expertise by such agencies.

Recent and anticipated changes to environmental legislation has also initiated a broadening of the scope of environmental impact assessments as well as the type of projects which require assessment:

- Under the AEPEA and the CEAA, the definition of environment has been broadened to include not only air, land and water (as under the current Alberta Environmental Impact Assessment Guide-lines), but all layers of the atmosphere, and all organic and inorganic matter and living organisms, and the interacting natural system that includes all these components. Under the new AEPEA, the potential impacts of a development from the environmental, social, economic and cultural perspectives must also be evaluated. Proponents could face delays in the review process and consequent in project development if all relevant issues are not addressed in the assessment document.

- The AEPEA, CEAA and Canada's Green Plan recommend EIAs take the ecosystem approach to resource impact evaluation where the interrelationships of biophysical components, as well as the biophysical components the themselves are discussed. This type of approach to environmental assessment preparation requires the integration of data from all disciplines and intense interaction of technical experts to delineate ecosystem relationships and subsequently define impacts.
- The CEAA requests that environmental effects of malfunctions or accidents associated with a project be reviewed. This requirement could have major implications for industry as potential events such as the failure of a tailings dyke, or a natural gas fire at a sour gas plant would need to be described.
- The AEPEA requires the evaluation of project effects on health and safety.

The complexity of EIAs is anticipated to increase as we start to include biophysical, socio-economic, health and safety risk issues, aesthetics, ecosystem effects, off-site effects, atmospheric effects, and the interactions of living organisms with organic and inorganic matter. The routine application of risk assessments will likely expand beyond examining the effects of water and soil contamination or sour gas blowouts on human health, to the examination of health effects on wildlife, livestock and forest resources. Under the new AEPEA and CEAA, the EIA must also address the effects of alternatives to the proposal, including the no go option, with respect to social, biophysical, economic and cultural issues.

Preparation of an EIA for a large or medium project is currently almost impossible to accomplish by traditional manual means. In the future, as the number of required EIAs increases and their complexity also increases, the preparation of a single environmental assessment, in a reasonable amount of time and at a reasonable cost, which adequately addresses, integrates and analyses all information and interactions of biological and physical, as well as social, health and economic consequences of development is going to severely stress the availability of experienced environmental impact experts. Something like a standardized EIA /GIS packaged will have to be used. And the requirement for contingency planning could be handily dealt with by developing the emergency response plans on top of the GIS-based information management system. Such a system would incorporate relevant predictive models to handle emissions of contaminants into the atmosphere, water or onto soil. It is also likely that the "look" of the EIA document will change, as large documents with bulky pockets for fold out maps are replaced by streamlined summary documents and references to digital maps in the GIS.

Properly designed, a GIS loaded with the baseline data can model interrelationships of biophysical components more quickly and reliably than if technical experts have to rely on traditional manual interaction.

## Off-site effects

A very important question, that has received greater attention over the past few years and has astronomical implications for the content of EIAs, is "Should off-site project effects be evaluated?" For example, if we are assessing the development of a sour gas processing plant or an oil sands project that will draw electricity from the Alberta grid, should the $SO_2$ emissions produced off-site from coal-fired facilities be addressed? Should sulphur dust deposited in Japan from solid sulphur transported from Alberta be addressed in the EIA?

In part, the question of whether to evaluate off site effects is moot. The evolution is definitely towards the discussion of environmental impacts on regional, national and even global scales. Recent involvement of the authors in public meetings related to major projects (e.g., the Caroline gas field and the OSLO project) clearly demonstrated that key stake holders are concerned about not only their own backyards, but also those of their neighbours. Under the Land Surface Conservation and Reclamation Act (LSCRA) and Alberta Environment's Impact Assessment Guide-lines, the Minister of Environment can request this information.

The implications, for industry, of this expansion in the scope and content of EIAs in addition to the increase in the number of types of projects which require them, will be an increase in the amount of effort and costs required to stay current on environmental issues, maintain up to date environmental databases and to objectively analyse that information in the preparation of assessment documents.

If we are going to develop sophisticated information management systems, national and international environmental data standards will have to be developed immediately; even, if to the chagrin of academics, they are *de facto* industry standards. Such information could be made available in public data libraries similar to the ones like the Nature Conservancy. We recognize that this is not a perfect solution but it is better than navel gazing until the "ultimate" format is developed.

## Preparing a focused, comprehensive and technically defensible EIA

Progress is accelerating towards the perfection of focused, comprehensive and technically-defensible evaluations of project impacts. An EIA that is incomplete or unreadable does not provide regulators or members of the public with the facts and analyses that are required for the project approval process. A satisfactory EIA should be succinct, well organized and clear (Ross 1987).

It must also be based on reliable, accurate, precise data assembled through the application of scientifically defensible (i.e., repeatable) methodologies.

The trend in the approach to EIAs is to prepare a focused assessment that addresses only key impacts to key environmental, social, economic and cultural resources, and to thoroughly document the analytical methods and decision-making process used to evaluate project effects through impact hypotheses. The AEPEA and ERCB support an environmental issues scoping process in Alberta, and at the federal levels, CEARC strongly advocates this process. In order to prepare a comprehensive and technically-defensible EIA, adequate baseline information on biophysical resources that could be affected by the project must be collected and a well-defined methodology to identify impacts and their significance, to quantify impacts and to model cumulative impacts must be utilized.

Many environmental assessments do not provide adequate information on the methods and assumptions used to evaluate impacts. Such "scientifically" incompetent studies limits the ability of public reviewers or regulators to assess the accuracy, reliability and comprehensiveness of impact predictions; that is, the decision-making process has not been thoroughly documented or justified. In some cases, these assumptions have accidentally been left unidentified or due to the inexperience of the author. In some cases there is inadequate baseline information upon which to make an assessment,and in still other cases, the proponents have strategically failed to document the information.

The use of a GIS application built according to a standardized design plan would provide something to audit. A review of such a GIS would quickly reveal whether the baseline information was complete and/or adequate. Deficiencies could be detected early.

## Regional biophysical baseline information

To effectively identify impacts, a thorough knowledge of biophysical resources, not only in the development area, but in the region being effected by project-induced impacts, is required. Baseline information used in assessments is usually inadequate for regional studies because it has not been collected over a broad enough region or over a long enough time period, or it has not been collected for all the biophysical resources or ecological components that should be examined in the EIA.

A confrontational question that needs to resolved soon, particularly in areas where several developments are being proposed (e.g., Bow Valley Corridor, Kananaskis Country, the Rocky Mountain House/Caroline area), is who should be responsible for collecting the regional baseline information that is required to satisfactorily evaluate project-induced impacts? One proponent is usually unable or unwilling to provide the money, time or effort necessary to collect baseline information outside their immediate project area. Unfortunately the provincial government, which is the steward and manager of most biophysical resources, also lacks funding for these regional programs. The most feasible approach appears to be the establishment of a collective of the developers, which is chaired by the government, and jointly designs and funds regional baseline inventories. In the Alberta Oil Sand Environmental Research Program (AOSERP) area in northeastern Alberta, information on biophysical resources was collected by the provincial government and made available to multiple users in the region. The AOSERP data, however, lacked a consistent format. A public "fee for service" environmental data bank of GIS compatible digital information could be a solution.

Use of a GIS encourages data collection and sharing. It would also provide a framework in which to use them. Data would be gathered according to selected criteria using objectively selected areal extents, density of sampling, etc. Such a request by a regulatory body is not so far fetched. Consider the fact that all the pages presented at this conference were submitted on diskette using a technology (desktop publishing) that was not recognized in business 10 years ago.

## Quantifying impacts and addressing cumulative impacts

A trend in impact assessment that will continue has been the attempt to quantify the effects of project activities on biophysical resources, and move away from subjective and qualitative impacts evaluations. GIS analysis has enormous potential to provide objective and defensible methods of quantifying spatial impacts.

Although advances have been made towards quantifying the magnitude, duration and spatial extent of impacts, these tasks have suffered from insufficient baseline information on biophysical resources. Progress in modelling research, and technical monitoring, have resulted in improved capabilities for predicting potential environmental impacts particularly in the area of water quality, hydrology, air and wildlife habitat. However, up to the present, the use of remote sensing information from airborne or satellite surveillance has be largely ignored in the preparation of EIAs. Major advances are still needed in the case of modelling effects of emissions, soil, vegetation,water, etc.

The evaluation of cumulative project-induced effects in EIAs will be required under both the AEPEA and CEAA. It is difficult to imagine how evaluation of cumulative effect would be possible without the use of GIS or something like it to facilitate modelling and quantification of impacts on the basis of remote sensing data and public libraries of digital environmental information.

Long-term monitoring and research programs provide information by which to verify the accuracy of predicted impacts and the effectiveness of proposed measures to mitigate adverse environmental effects, and thereby improve the ability to predict real impacts. GIS offers the mechanism by which research and monitoring data could be systematically and routinely compared with baseline conditions or predictions. It is likely that discussion will finally determine that data reported to the public regulatory agencies will be made available to the public on a discretionary basis. Research data, however, are likely to remain proprietary.

## The future

Under the proposed provincial and federal acts, both the mandatory and discretionary approach will be taken in determining whether an activity will require an environmental assessment. The new AEPEA will include a list of projects for which an EIA will be mandatory, and the Director and Minister of Environment will have the discretionary powers to request an assessment if the activity has significant adverse impacts, the impacts of the project are unknown, or there is major public concern.

In the future, we anticipate that significant changes in the EIA process, as such assessments become more common and more objective. Historically, EIAs have been prepared by the project proponent and technical experts, under government guidance and recommendations, and public involvement has been restricted to hearings. However, this approach may change in the future because proponents can lack objectivity and may be biased. And are government officials are independent of political influence? The Environmental Legislation Review Panel to Alberta Environment on AEPEA recommended that "a third party prepare an EIA report in circumstances where legitimacy and public acceptance of a proponent-prepared report might be in doubt". A committee, appointed by the Saskatchewan Minister of Environment and Public Safety to review the province's environmental assessment legislation and practice, presented its recommendations in February 1991. The committee stated that "those responsible for the EIA must be objective, unbiased and independent of political influence". We suggest that professional environmental experts who are controlled by an enforced code of ethics and practice (e.g., A.S.P.B.) using GIS according to strict protocols will provide an environmental impact assessment that is the objective, unbiased and free of political influence. The protocols are expected to be set by a regulatory agency similar to the NRCB in Alberta.

It must be obvious that the days of experts working in isolation are disappearing. The use of a interdisciplinary networked GIS-based information systems are starting to provide models of the real world. This should improve the understanding and reliability of predicted impact assessments, and at the same time resulting in cost reductions by eliminating (or at least reducing) the collection of redundant, incompatible or contradictory data.

In summary, these changes are going to result in an increase in the number of EIAs being prepared. In order to facilitate the prepared preparation and review of EIAs in a timely manner, it would be beneficial if they were implemented on competent and compatible GIS platforms according to carefully defined designed considerations. It is anticipated that regulatory agencies will soon begin issuing recommended (or required) GIS design considerations to be followed in the preparation of an EIA.

It is apparent to us that the GIS has the potential to streamline and standardize the process. Geographical units can be mapped and ranked values assigned for important criteria. Also, the criteria can be weighted to reflect their perceived importance. Boolean algebra can then be used to identify the key issues. One area where this type of quantitative and objective approach is very important is in site and route selection. The use predictive models (integrated into to GISs) is another area which could contribute greatly to the conservation of time, money and paper by identifying those "significant" impacts so they could be incorporated into an environmental information system.

Remote sensing (LandSat, SPOT, radar, microwave, airborne multispectral, etc.) needs to be taken advantage of to provide the regional background information necessary to allow off-site and cumulative impact evaluation.

Design of GIS-based environmental management information systems needs to be researched with the view of developing a practical industrial design for such an application. We believe that this will and should be an evolutionary process leading to the refinement of a minimal GIS-based system which facilitate environmental management throughout all phases of a project; planning, approval, operation, emergency response, decommissioning and reclamation. Such a system would be eminently suitable for adaptation for use in developing countries. It is vital that the evolutionary process maintain upward data compatibility.

Public libraries of digital environmental information generated by EIAs need to be planned and implemented by an unbiased party. A good model is the ERCB well data library in Alberta. A good candidate to operate the library for Alberta, is the NRCB.

## References and other background materials

Alberta Minister of Environment. 1991. Bill 53 - Environmental Protection and Enhancement Act. 127 pp.

Beansland, G.E. and P.N. Duinker. 1983. An ecological framework for Environmental Impact Assessment in Canada. Published by Institute for Resource and Environmental Studies, Dalhousie University and Federal Environmental Assessment Review Office. 132 pp.

Beitz, B.F. 1991. Environmental mandate of the ERCB. pp 49-51. *In:* M.H. Hamza (ed). Proceedings of the IASTED International Symposium World Environment. Held in Calgary, Alberta, April 8 - 10, 1991.

Clean Air Strategy. 1991. Clean Air Strategy for Alberta: Draft Framework Document. 135 pp.

Duffy, P.J.B.(ed). 1986. Initial Assessment Guide Federal Environmental Assessment and Review Process. FEARO, Ottawa, Ontario.

Federal Activities Branch, Environmental Protection Service and FEARO. 1978. Guide for Environmental Screening. 84 pp.

Federal Minister of Environment. 1991. Bill C-13 - An Act to establish a Federal Environmental Assessment Process.. 41 pp.

Government of Canada. 1979. Revised guide to the federal Environmental Assessment and Review Process. 12 pp.

Leopold, L.B., F.E. Clarke, B.B. Hanshaw and J.R. Shaw. 1971. A procedure for evaluating environmental impacts. US Geological Survey Circular 645. US Geological Survey, Washington DC. 13 pp.

Marshall, D., W.I. William, W.A. Ross and B. Saddler. 1986. On scoping in EARP. Federal Environmental Assessment Review Office, Vancouver, B.C.

Mitchell, R.H.(compiler). 1991. Clean Air Strategy for Alberta: Background Project Reports.

Praxis. 1988. Public Involvement: Planning and Implementing Public Involvement Programs: Executive Overview. Prepared for Federal Environmental Assessment Review. 119 pp.

Ross, M. 1991. Perspectives on joint environmental impact assessments. *In:* M.H. Hamza (ed). Proceedings of the IASTED International Symposium World Environment. Held in Calgary, Alberta, April 8 - 9, 1991.

Ross, W.A. 1987. Evaluating environmental impact statements. J. Envir. Manage 25: 137-147.

Sacks, A. and P.B. Clark. 1980. Scoping the content of environmental impact statements: an evaluation of agencies' experience. Draft report for Council on Environmental Quality and Geological Survey, US Department of the Interior. 92 pp.

Wolfe, L.D.S. 1987. Methods for Scoping Environmental Impact Assessments: A Review of Literature and Experience. Prepared for the Federal Environmental Assessment Review Office, Vancouver by Larry Wolfe Associates. 50 pp and Appendices.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Monitoring and modelling transboundary air quality using GIS

**Anne Lucas**

IBM-Bergen Scientific Centre
Høyteknologisenteret
Thormohlensgate 55
N-5008 Bergen, Norway
Tel: (+47-5) 54-46-15 Fax: (+47-5) 96-01-15

## Abstract

*A GIS is used to integrate geographical, meteorological and chemical data with model results furnished by a number of long-range atmospheric transport models. By combining the results from a series of model simulations, new insight into the spatial and temporal variability of predicted air quality can be gained. Several pollutants, including arsenic grid established by EMEP (Co-operative Programme for Monitoring and Evaluation of the Long-Range Transmission of Air Pollutants in Europe), covering Europe at a 150 km interval.*

*Overlays are created for individual and combined model input parameters (winds, precipitations, mixing height, scavenging ratio). Monthly and annual patterns of emission, deposition (wet/dry) and concentration are described for single chemicals and chemical combinations. These are further combined with geographic variables such as soil characteristics (slope, texture, cultivation index), vegetation and topography, for impact assessment.*

# C4 GIS in information analysis

**Fast access to fisheries information with low-cost desktop mapping software**

Peter Lewis et al., Ministry of Environment, Lands and Parks, Victoria, BC, Canada

**The conceptual determination of carrying capacity analysis using GIS technology. Case study: the upper Komering sub-watershed, South Sumatera, Indonesia**

Nanna Suryana et al., The Wageningen Agricultural University, Wageningen, Netherlands

**Setting global and regional priorities for coastal temperate rain forest conservation**

Andrew Mitchell, Conservation International, Washington, DC, USA

**Evaluating alternative land management policies from a water quality perspective: an automated land information systems approach**

Lucy Buchan et al., University of Wisconsin, Madison, WI, USA

GIS'92

Symposium

Vancouver
British Columbia
February 1992

# Fast access to fisheries information with low-cost desktop mapping software

**C. Peter Lewis and**
**R.S. Hawthorn**

Fisheries Branch
B.C. Ministry of Environment, Lands and Parks
780 Blanshard Street
Victoria, B.C.
Canada V8V 1X5

## Abstract

*The most urgent problem currently facing agencies like ours is our inability to rapidly locate existing information, some digital, but most still hard copy in files, map cabinets, etc. Information about the aquatic resources of streams and lakes in British Columbia has been collected for more than 40 years but it is scattered in various offices throughout the Province and many potential users are not even aware of its existence. The challenge is to provide easy low-cost access to that information, including access by geographic location.*

*This paper will illustrate the system that was developed to enter, store, map, query and retrieve geo-referenced information on the fisheries resources of B.C. The approach taken was incremental rather than all encompassing and is described in detail. Design and development were carried out by existing staff (biologists and technicians) using QUIKMap, a locally developed desktop mapping package that retails for less than $1,000. Internal data are integrated with other relevant digital data provided at many different scales and from many different sources. Screen and printed reports include maps, tables and text. The Ministry of Environment's 1:50,000 digital Stream Atlas for the Province is used as the base.*

## Introduction

*What's This All About?* For the past two years, provincial (Fisheries Branch, B.C. Ministry of Environment, Lands and Parks) and federal (Habitat Management Division, Dept. of Fisheries and Oceans, Pacific Region) fisheries organizations have been experimenting with desktop mapping, called "simple" or "non-analytical" GIS by some, as a practical means of providing fast, low-cost access by geographic location to existing fish, fish habitat and related administrative data.

In many operational situations, fisheries field staff function not only with little hard information about their resource but with inefficient and incomplete access to that information which does exist. This paper describes our experiences with simple GIS as one potential remedy for the data access problem and includes:

- reasons for beginning with desktop mapping — the bite-size approach;
- the process for defining and marketing the system;
- administrative and resource themes for fisheries management;
- options and problems in selecting a digital base;
- capabilities and limitations of the desktop mapping software package, QUIKMap by Axys Software Ltd. of Sidney, B.C. (formerly ESL Environmental Sciences Ltd.); and
- where we think we're at and where we're going.

## Choking on GIS — is a seven-course meal the only way to go?

Conventional wisdom suggests that the adoption of GIS technology by an organization can be successful only if carefully planned and fully resourced. Design and implementation should be top-down, preferably driven by a broad range of applications, controlled by a fully specified systems development process, and carefully piloted (Campbell, 1989; Prior, 1991 etc.). This is seven-course GIS. The proceedings of GIS conferences like this are awash with examples: Bell (1991); Johnstone et al (1991); Zillmer (1991) to list a few. We note with some suspicion, however, that all too commonly, presentations focus on system specification and design and, all too rarely, include testimonials from line managers and staff about the ongoing effectiveness of GIS in their daily work.

Fisheries Branch and DFO Habitat have discussed GIS ad nauseam for at least the last 10 years, even before many knew what the letters stood for. Meetings, workshops and JAD's seem to occur monthly; memos and notes circulate back and forth, filling both file drawers and hard disks; strategic implementation plans have been started and even finished; budget proposals are prepared and submitted and revised and resubmitted; broader government initiatives have been supported and depended upon — and, for all this, we are not much closer today to making effective use of our georeferenced information than we were 10 years ago.

By traditional criteria, our organizations are not ready for GIS. Corporate sponsorship has been soft, potential in-house project managers and trained staff are almost non-existent, the limited number of digital data bases that we do have are not properly georeferenced, operational staff are frequently not computer-literate much less GIS-literate, and funds have been in very short supply. But the demand exists and is growing rapidly, fueled more by need and propaganda than by rational expectation. Both senior management and field staff have been infected but remain suspicious. The few individuals who do have more than an introductory understanding of GIS and of database management, both users and systems professionals, have been trapped in the middle.

Within this context, a few of us have sought an approach to fisheries GIS implementation that would:

- be incremental rather than all encompassing,
- be practical given our resources and expertise,
- permit initial focus on database creation, and
- both seduce and educate staff with simple applications of immediate use.

We like to call it bite-size GIS (more a wish than an expectation) and our hope for it springs both from the apparent failure, at least in the short term, of seven-course GIS in many organizations far more able than ours, and from the bottom-line fact that we are simply not capable of any more at the present time.

## Grab the ball and run with it

Bite-size or not bite-size, we would have felt more comfortable following an explicit and accepted systems development process for this initiative. This was just not possible, except in an overview sense. We did not have either the funding or the organizational support. We had to create products and demonstrate their value for unsophisticated users before GIS of any sort could expect the support and resources necessary for success in our organizations. Everyone was tired of talking about GIS. If we were going to do anything (and we accept that we probably shouldn't have), we had to grab the ball and run with it and we had to use our own staff to do so.

Our approach was to rely initially on the expertise and insight of a few individuals, both in-house and associated with our software vendor, and to assume that the technology we could afford was the technology we needed (fortunately, we really believed that it was). The dangers in this are well known — adoption of dead-end technology, creation of databases that don't meet broad standards and must be abandoned or expensively modified, inability to meet user needs leading to frustration and abandonment, closing off of future options, etc. The horror stories are never ending. We were aware of these but hoped that the awareness itself, together with the narrow focus of what we were trying to do, would minimize the risk to both ourselves and our organizations.

The entire effort would not have been possible without a fortunate combination of circumstances and capabilities that included:

- an existing 1:50 000 digital Stream Atlas for the entire province, created internally and available free of charge, that could be used as a base;
- a piece of desktop mapping software that was affordable and well supported by a company that was willing to work very closely with its customers;
- a computerized Stream Information Summary System (SISS) (Pontus and Komori, 1990) whose use was severely limited by reliance on associated manually-generated maps that could not be easily updated;
- several very experienced staff and an excellent Fish Habitat Inventory and Information Program User Needs Analysis (Mathers et al, 1985) that gave us initial confidence in the value of what we were trying to do; and
- last, and critically, a temporary position that we were able to fill with an inventory biologist who had gone back to school to learn about computers.

We began by creating a computerized version of a single SISS map, enhanced to include lake as well as stream data, using the Stream Atlas file for a local 1:50 000 map sheet area on Vancouver Island (NTS 92F/14: Oyster River). The SISS data were selectively augmented with information from other digital base maps, digitized administrative boundaries, and data imported from existing fisheries databases. The process was painful. Even our internal databases were not compatible with each other: rainbow trout, for example, were coded RB in one database and RT in another; local names were used in one and official names in another; values for an attribute in one that directly contradicted those in another; and so on. Special symbols had to be designed and procedures developed for placing those symbols in cartographically appropriate ways on the map. Despite our focus on data access and display rather than analysis, we continually pushed the software to and beyond its limits. However we found that we could usually do what we wanted, either by breaking rules or by convincing the vendor to modify the software. In the end, we were able to display and plot both map and tabular products that we believed would be of value and interest to our operational staff.

## We didn't forget our audience

The next step was to see if anyone agreed with us. We took our computer to several internal meetings involving both headquarters and field staff, discussed our efforts, and showed them the products for the Oyster River map sheet. For many people, this was the first time they had actually seen GIS, either simple or full-featured, using real data that was of relevance to fisheries management. The reaction was enthusiastic, not just for the products, but for the fact that we had done them in-house, using software that seemed affordable and relatively easy to use. At the same time, we started to get the feedback that was essential if we were to improve the product set.

This need for feedback, coupled with our hidden agenda of using staff familiarization and involvement to generate demand for GIS within our organizations, led to a much more ambitious "pilot" initiative. We duplicated the Oyster River product for one 1:50 000 map sheet area in each of the 10 B.C. Fisheries sub-regions, covering all DFO Divisions in the process, and hauled our computer by road (we didn't trust air transport) to each office in order to present and discuss both the local map sheet and regional data. Our presentations were directed primarily at Fisheries and Habitat staff but the audience included interested personnel from other programs and agencies. In all, 149 people attended the regional demonstrations and many more from a broad spectrum of organizations requested and received demonstrations in Victoria — and continue to do so.

In each case, the system was not presented as a *fait accompli*. Our visit was advertised as an opportunity to use real products as the basis for consideration of regional needs and priorities for geographic information, particularly as they related to the objectives of the Fisheries program but also from the general point of view of the umbrella organizations. Discussion focussed on a set of questions distributed at the beginning of each presentation and encompassed:

- general regional issues and concerns;
- relative priority of new inventory versus improved access to existing information;
- simple data display versus complex spatial analysis;
- digital base map requirements;
- overall GIS strategy, action plan and resourcing;
- additional data displays that would be of use;
- issues related to the creation and management of digital databases; and
- the role of headquarters versus regions.

As a result, we not only received productive feedback about specific themes and capabilities that would be of value in the management of fish and fish habitat, but were able to brainstorm about directions that would be most appropriate for broad GIS implementation in the longer term within our organizations.

## Themes for fisheries management

The hardcopy Stream Information Summary Map that we began with contains information about:

- hierarchical watershed codes;
- fish distributions and spawning areas;
- the locations and nature of blockages and barriers;
- the locations of other "points of interest", e.g. hatcheries, spawning channels, flow gauging stations.

We also knew, from a provincial government point-of-view at least, that lakes were as important as streams. At a minimum, field staff need to know which lakes have been surveyed and when, what data is available, where and in what form. Beyond this, our travels confirmed that administrative themes are as important to habitat and fisheries managers as information about the natural resource itself. Management unit boundaries, stocking information, waterbodies with special angling regulations, guide territories, angler use survey locations — all were included in the many types of georeferenced information that staff believe they need to access.

The QUIKMap products that we prepared to address these requirements are listed in Table 1. Four of the demonstration screens, showing both administrative and resource data, are reproduced in Figure 1. Figure 1a identifies lakes that have been completely or partially

Figure 1. Fisheries Summary Information: Sample Screen Plots

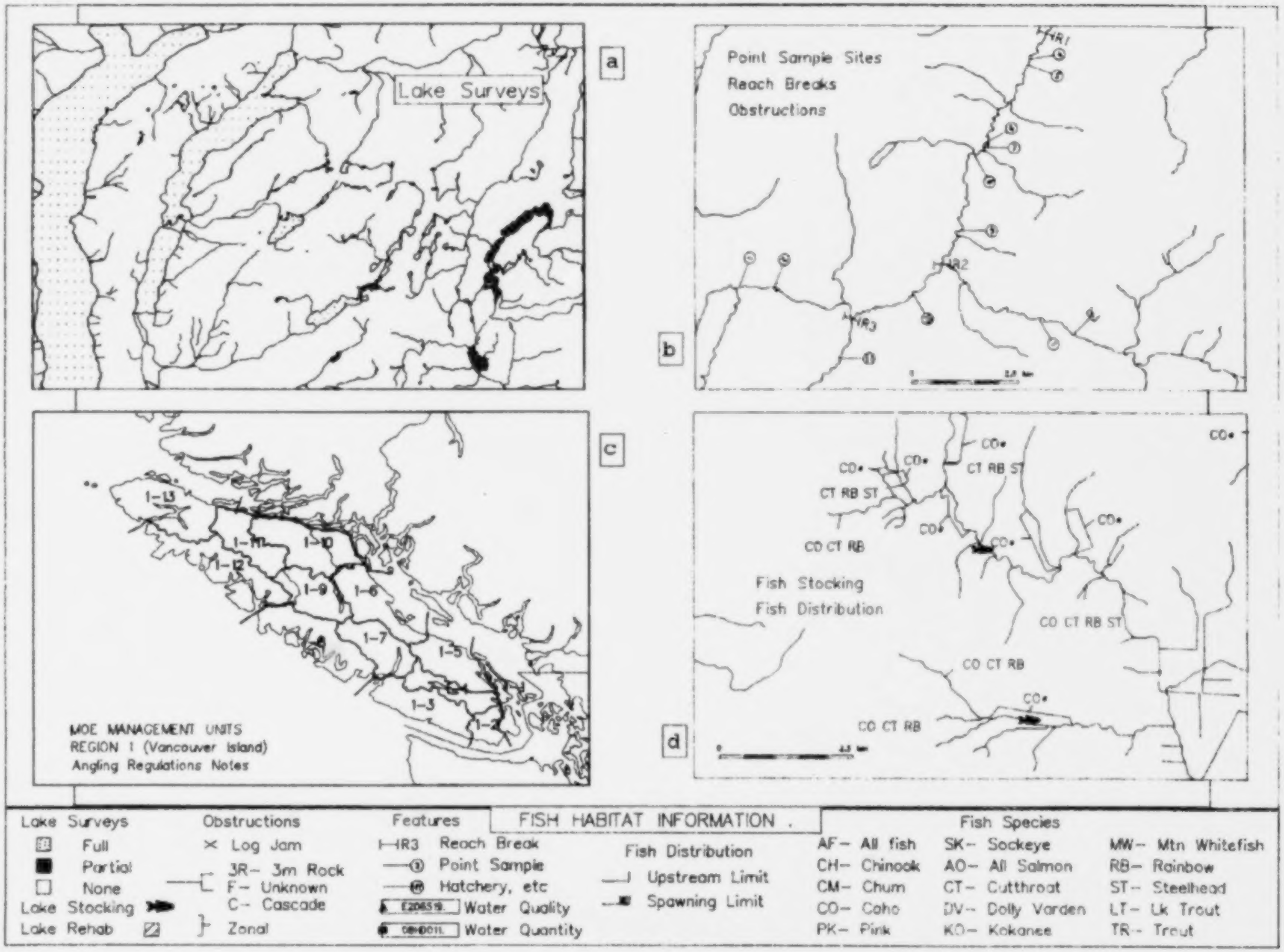

a
Lake Surveys
b
Point Sample Sites
Reach Breaks
Obstructions
R1
R2
R3
c
1-13
1-11
1-10
1-12
1-9
1-6
1-7
1-5
1-3
1-2
MOE MANAGEMENT UNITS
REGION 1 (Vancouver Island)
Angling Regulations Notes
d
CO•
CT RB ST
CO CT RB
Fish Stocking
Fish Distribution
CO CT RB ST
FISH HABITAT INFORMATION .
Lake Surveys
Full
Partial
None
Lake Stocking
Lake Rehab
Obstructions
× Log Jam
3R– 3m Rock
F– Unknown
C– Cascade
Zonal
Features
R3 Reach Break
Point Sample
Hatchery, etc
E206519 Water Quality
08HD011 Water Quantity
Fish Distribution
Upstream Limit
Spawning Limit
Fish Species
AF– All fish
CH– Chinook
CM– Chum
CO– Coho
PK– Pink
SK– Sockeye
AO– All Salmon
CT– Cutthroat
DV– Dolly Varden
KO– Kokanee
MW– Mtn Whitefish
RB– Rainbow
ST– Steelhead
LT– Lk Trout
TR– Trout

**Table 1. QUIKMap Product Set for Fisheries Management.**

### a. Provincial

- 1:2,000,000 overview: coastline, islands, major water features, NTS 1:250,000 grid, local map areas.

### b. Regional

- sample base maps at scales of 1:20,000 to 1:525,000 combined on screen: TRIM, forest cover, federal 1:250,000, hydrographic.
- watershed boundaries for major systems.
- administrative boundaries: MOELP Sub-regions and Management Units (MU's), DFO Sub-districts.
- data displays tied to administrative areas: angling regulations notes, stocking records.

### c. Local 1:50,000 Map Areas

- fish distribution and spawning areas: summary and by individual species.
- blockages and barriers.
- points of interest: hatchery, counting fence, etc.
- point and reach survey sites; summary of available information.
- water quantity and quality sample sites.
- critical habitat areas.
- surveyed lakes; summary of available information.
- stocked streams and lakes; accompanying data tables.
- angling guide territories.
- aerial surveys of angler use of lakes.
- watershed codes.
- watershed boundaries for individual streams.
- biogeoclimatic zones.

surveyed. Clicking (with a mouse) on a surveyed lake will bring up a summary and description of the survey data. A scanned bathymetric map can also be imported. Figure 1b displays reach breaks and point sample sites from detailed aquatic biophysical surveys, together with the locations of obstructions to fish passage. Angling regulations notes can be listed by clicking on a management unit in Figure 1c and Figure 1d shows both fish distribution by species and stocking locations. Major legend groupings are colour-coded so that multiple themes can more effectively be reproduced on one map — all obstructions are in red and all fish species codes in green, for example.

For the most part, this is overview-level information. Potential management applications include:

- identification of high value and threatened wild fish stocks;
- location of critical habitats and controls;
- broad estimates of productive capacity;
- summary assessment of existing and potential impacts.
- selection of sites for enhancement or restoration activities; and
- development and modification of regulations.

And, even at this level, an average of four staff days was required to add the available thematic information to each 1:50 000 base. Stream Information Summary (SISS) data sets are not properly georeferenced and couldn't be migrated directly. Other digital data are stored on a variety of hardware/software platforms, usually in a manner that makes sharing difficult or impossible, this in addition to the inconsistencies in content already discussed. And most data is hardcopy, scattered throughout various regional and headquarters offices, much of it unknown to the staff that need it most. The collation process was a graphic illustration of the difficulties that field staff must have in attempting to gather information in order to respond to operational demands.

## Base wars — streams are different!

The MOELP digital Stream Atlas has been essential to this initiative. The Atlas is a data base of the complete British Columbia stream network ("bluelines") shown on 1:50 000 NTS map sheets except for some high gradient/low fisheries value 1st order tributaries. It was created to provincial standards on a VAX minicomputer using Intergraph IGDS/DMRS and custom software developed by PAMAP Technologies. Most importantly, the Atlas is not simply a base map upon which data can be overlain. All streams are uniquely identified by a computer-generated watershed code that locates it in the watershed's tributary hierarchy and to which any data can be tied. Associated attribute data include:

- gazetted stream/lake names;
- 1:50 000 NTS sheet number for watershed mouth or lake outlet;
- UTM grid reference of mouth or outlet;
- distance upstream from trunk mouth (in m);
- lake sequence number; and
- lake inlet and outlet locations.

At the present time, the Stream Atlas is, in the view of both provincial Fisheries and DFO, the only province-wide base with sufficient — but not excessive - detail and structuring to be effectively used with watershed-level fisheries data, the highest immediate priority for our inventory programs and, as we have seen, the level toward which this pilot initiative was focussed. Other possibilities are listed in Table 2, along with the advantages and disadvantages of each from our point-of-view. All, including the Stream Atlas, have significant shortcomings, either in content, in quality or in coverage. Even the Atlas, while essential, was not viewed by operational staff as enough in itself. Background topography, toponymy, transportion routes and settlement locations, at a minimum, were also seen as important for convenience and effectiveness of use. This could be accomplished in QUIKMap by combining the Atlas with selected themes (layers) from the federal 1:250 000 digital coverage — also immediately available for the entire province.

Unfortunately, the Stream Atlas has been temporarily mothballed because of disagreement over whether it should or should not be part of a supported provincial base. The Provincial Digital Atlas is based on three distinct data sets with levels of detail appropriate for display at 1:2 000 000, 1:250 000 and the TRIM product at 1:20 000. Some individuals see the Stream Atlas as competitive with TRIM, the future for detailed applications, but, for our immediate overview purposes, we see the Atlas only as a necessary enhancement to the provincial 1:250 000 base (itself an extension of the existing federal product; see Table 2). For aquatic information, streams are different. The streamline — or segment of it — is not merely a backdrop against which data is overlain; stream segments are the fundamental spatial units to which that data is tied. Either we have a base that includes the level of detail and the features of the Stream Atlas or we must "turn off" the existing streamlines on that base and overlay the Atlas.

Streamlines on the provincial 1:250 000 base are unacceptable because of highly varying density and quality from map sheet to map sheet (because of different data sources) and because of insufficient structuring — no watershed code or equivalent. As well, lake outlines are too generalized, even for overview aquatic requirements, and smaller lakes are not captured at all. Watershed codes are very important in our work, primarily for external communication. The code provides a standard identifier for tying attribute data to stream graphics, a convenient and unique label for hardcopy filing systems, and a way of ensuring that all parties are talking about the same piece of landscape in planning or problem-solving exercises. The system is described in detail in "A Guide to the Hierarchical Watershed Coding System for British Columbia" (British Columbia, Ministry of Environment and Parks, 1988).

So the digital Stream Atlas remains in IGDS/DMRS format, a format that is no longer supported in MOELP. Fortunately, both for this initiative and for the rapidly increasing number of other users, within and outside government, it is usable. We can convert to QUIKMap and other GIS formats. We do not, however, have the ability, in-house, to edit it or to add new streams and watershed codes. Nor can we properly enhance it with lake outlines and watershed boundaries, the two highest priorities for the future, although both were added for the map sheets used in this project.

**Table 2. Digital Base Maps for B.C. Streams and Lakes**

| Base | Advantages/Disadvantages |
|---|---|
| MOELP Stream Atlas<br>1:50,000 | Advantages: complete provincial coverage; consistent standards to provincial level including edge matching; unique stream identifiers provide intelligence.<br>Disadvantages: streamlines only; lake inlet/outlet locations but no outlines; no cultural or topographic information; presently archived; cannot be maintained or updated. |
| MOF Forest Inventory<br>1:20,000 | Advantages: includes lake outlines, lake and stream names, road network, other cultural and resource data.<br>Disadvantages: doesn't cover non-forest areas (parks, alpine, urban centres); streamlines break at bridges, map boundaries; CAD format; no network intelligence; quality control problems. |
| Surveys & Mapping<br>DEMR, Ottawa<br>1:250,000 | Advantages: complete provincial coverage; includes lake outlines, generalized cultural and topographic information including toponymy.<br>Disadvantages: inconsistent level of detail for streamlines; lake outlines generalized and smaller lakes not captured; no intelligence; quality control problems. |
| MOELP, Surveys & Resource Mapping<br>1:250,000<br>(enhanced) | Advantages: as for federal coverage above except coverage not complete until 1993; format consistent with TRIM; additional editing; digital elevation model.<br>Disadvantages: as for federal coverage above except much improved quality control. |
| MOELP TRIM<br>1:20,000 | Advantages: includes lake outlines; detailed cultural information including toponymy; detailed DEM; high quality; meets all federal and provincial specifications.<br>Disadvantages: only 1/3rd of the province has coverage; completion presently scheduled for 1998; excessive detail and file size for overview applications. |

## Desktop mapping — is it enough?

The short answer is no, but for many immediate purposes in fisheries management simple GIS will do, at least for display if not for data entry, until something better comes along. The highest priority demand in most offices was for quick and easy access to up-to-date basic information in order to reduce the burden of responding to routine questions and issues that are largely administrative. The need for complex spatial analysis was recognized but many staff felt that simple map/data display on many desks in each office would meet 80-90 per cent of their needs. Direct public access to menu-driven read-only workstations was suggested. An absolute requirement that software be easy to learn and operate was repeatedly emphasized. There is no faith at the working level in our organizations that full-scale GIS implementation with adequate technology and technical support will take place in the foreseeable future.

This feedback reinforced our belief that QUIKMap, or at least a QUIKMap-type product, would be a productive platform upon which to tie our initial forays into the hazardous world of GIS. QUIKMap's features are summarized in Table 3. The critical points to note are:

- the ability to display and plot anything on any base;
- low cost and limited hardware requirements for both QUIKMap and its associated database management software;
- the absence of analytical capability;
- the ability to import from a variety of full-featured GIS and CAD software; and
- the relatively easy learning curve.

A disadvantage for us, although not a fatal one, is that only a PC version of the software exists; Mac hardware is not supported.

A much more serious "fly in the ointment" and one that we remain very seriously concerned about is the use of any simple GIS package for digitizing line and polygon data. Problems with data display are one thing, software-generated limitations or errors in data acquisition quite another. QUIKMap, for example, does not by default store digitized data in a way that supports topological intelligence. This was not a significant factor for our pilot work but could be for full provincial implementation. A sudden plunge into an extensive data entry program without a full understanding of this issue would be a classic example of the danger in being technology-driven.

## Where did we get with all this? What next?

The most realistic perspective is probably the statement, "12 down, 1162 to go!" There are 1174 1:50 000 NTS map sheets for British Columbia. We have created a product set for fish and fish habitat management that is valuable in the view of both line managers and staff but we are not much closer to full implementation than most of the seven-course meal GIS projects. On the other hand, we have spent a lot less money and haven't made any mistakes that we can't easily afford to back away from. Bite-size GIS does have its advantages.

We have been forced to face up to some fundamental truths because of this exercise. The most basic one is that, no matter how you begin, GIS implementation is going to be data-controlled. The data itself will quickly become the main focus and the main investment in building a system. Yes, this statement is obligatory in any GIS paper: it is a truism and it is fast becoming trite. But it's one thing to read it and another to live it.

As organizations, we do not have vast amounts of data at hand: our operational staff spend as much time or more complaining about not having information as they do about not having access to the information we do have. Even for an overview-level GIS system as discussed here, though, we are looking at about 20 person-years of work to create properly georeferenced data sets for the Province. And this is on top of the 12 or so person years it took to create the digital Stream Atlas and the additional 5 plus person years that will be required to add lake outlines and watershed boundaries. From our point of view, this is "big business".

Still, this initiative has shown that there is a role for simple, inexpensive map display software in GIS for fisheries management. The concerns about data entry are real and must be taken seriously, but for data display and given their ease of use, products like QUIKMap can be extremely effective. We note with interest that the B.C. Ministry of Forests has backed away from reliance on full-featured GIS in its district offices, at least for the short term, largely because of data quality problems and lack of qualified staff. Instead, they have used QUIKMap and FoxGraph to create a Forest Inventory Reporting System (FIR) that provides access from districts to their central timber, range and recreation inventory database. Products are user-defined statistical tables, coloured charts and colour-themed maps. This is a bite-size approach by an organization that does have a large volume of digital data in place.

And we have created a monster! We now have a more informed and a more demanding clientele. Staff, even senior management, have been seduced and are looking for solutions, both short and long term. Increased high level support ensures that our own seven-course meal is coming — someday. We can look forward to LAN/WAN client-server UNIX based technology, data administrators and GIS technicians, properly specified databases full of quality data, the Land Information Infrastructure, TRIM and a host of other goodies. But our staff have already had a taste of the turkey and our history indicates that no one is going to resist snacking until the meal is served. Ease of use alone will generate use. We don't think that's a bad thing.

## Table 3. An Overview of QUIKMap.

- a simple but effective desktop mapping package that is very good at producing high quality output screens and plots.
- maps can be displayed in any of 6 different projections and the map display can be manipulated by windowing and enlarging specific areas (zooming).
- provides a scale bar which automatically re-sizes after zooming and the ability to display cursor coordinates in UTM or Lat/-Long.
- at any time, the active display can be printed, plotted, saved as a screen image or exported in .DXF format.
- not a true GIS by most definitions because of limited internal analytical capability: areas, line lengths, some neighbourhood operations and not much else.
- costs less than $1,000 at this writing, and can be used effectively on a standard VGA colour PC workstation (preferably 386SX or better) with a math coprocessor.
- present version uses dBASE III Plus (or a compatible like FoxBASE+) or Excel, and the next release will support dBASE IV.
- differs from most GIS systems in that data objects -- points, lines and polygons (areas) -- are not fundamentally linked to the base maps on which they are displayed; instead, objects are stored in data files and may be displayed (overlain) on any base map.
- custom base maps can be created by combining bases, regardless of initial scale, and by turning layers on or off as desired.
- information about an object is stored as a database record which can be displayed by pointing to it on the map or extracted by performing one of a number of possible queries in dBASE.
- includes a digitizing module which automatically calculates line length and polygon circumference and area.
- translators are available to permit conversion of spatial data captured in other common formats (including Autocad .DXF, Intergraph .DGN, CARIS .NTX, and TRIM).
- computer-literate staff can be functional in QUIKMap with only a few days training and do not have to work continuously with it in order to retain operational familiarity.

So what are the next bites? Briefly,

- Field offices are buying QUIKMap. A small angling guide database is already in place in one MOELP region. We intend to encourage and support limited initiatives like this.
- Many of our habitat staff think that the Forest Inventory Reporting System (FIR) would be extremely valuable for their own work. MOF have offered to install and train. We're going to take advantage of that offer.
- In headquarters, we are moving toward an accepted systems development approach and are concentrating on the specification and development of province-wide databases, both spatial and attribute, and on database management processes. At a minimum, we want to provide systems and standards for the collection and storage of stream and lake data that will be broadly used throughout the province. We realize that outside agencies *(native organizations and B.C. Hydro, for example)* are collecting more information than we are and we want to be able to make full use of that information and to be able to easily provide ours to them.
- Further implementation of the product set developed in this initiative — a Fisheries Information Summary System — depends on the existence of a functioning digital Stream/Lake Atlas. This is our highest priority and, because of our efforts and rising external demand, we finally have management support. Revised specifications for the Atlas, including conversion to an arc-node, 3-D network topology, are near completion and will be implemented on a full-featured GIS workstation by spring of this year. We have also begun to add lake outlines for priority areas and are looking at how we are going to link to the 1:20 000 TRIM base for our detailed reach-level data.
- Our biggest problem is to remain nearsighted enough to achieve concrete results — i.e. not to be seduced ourselves by the seven-course approach. At the same time, however, we have to avoid the dangers of being nearsighted — choosing dead-end directions and letting broader initiatives proceed by default along routes that conflict with our interests. For these reasons, we remain involved in and informed about the plans for seven-course technologies in our organizations. To the greatest extent possible, though, we are letting others worry about the details.

## References

Bell, Wm. C. (1991). Integrated Information Systems for Multiparticipant Projects. **Proceedings, The Canadian Conference on GIS-91.** March 1991. Ottawa, Canada. Abstract.

British Columbia, Ministry of Environment and Parks, Water Management Branch. **A Guide to the Hierarchical Watershed Coding System for British Columbia.** Victoria, B.C. 33p.

Campbell, G.J. (1989). Management Considerations in Applying GIS Technology within an Organization's Information System Strategy. **Proceedings of the 1989 National GIS Conference.** February/March 1989. Ottawa, Canada.

Johnstone, Wm. M., Hawkings, B. and Crooks, L. (1991). Development of a GIS Implementation Plan for the Yukon Department of Renewable Resources. **Proceedings of the GIS'91 Symposium.** February 1991. Vancouver, B.C.

Mathers, J.S., Chamberlin, T.W. and Delaney, P.W. (1985). **Federal/Provincial Fish Habitat Inventory & Information Program, User Needs Analysis.** Internal Final Report. Canada, Dept. of Fisheries and Oceans, Pacific Region and British Columbia, Ministry of Environment, Fisheries Branch. 40p. + appendices.

Pontus, K. and Komori, V. (1990). **Fish Habitat Inventory and Information Program, Stream Information Summary System (SISS) User's Manual.** Canada, Dept. of Fisheries and Oceans, Habitat Management Division. Vancouver, B.C.

Prior, N. (1991). Interpreting the "IS" in GIS. **Proceedings, The Canadian Conference on GIS-91.** March 1991. Ottawa, Canada. Abstract and Presentation Preprint.

Zillmer, G. (1991). Implementing GIS Technology: A Management Perspective. **Proceedings, The Canadian Conference on GIS-91.** March 1991. Ottawa, Canada.

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# The conceptual determination of carrying capacity analysis using GIS technology. Case study : the upper Komering sub-watershed, South Sumatera, Indonesia

**Nanna Suryana**

Department of Land Surveying and Remote Sensing
The Wageningen Agricultural University, Wageningen
The Netherlands

**Yyolanda Benitez Gomez**

GISRA, the Wageningen Agricultural University
Wageningen, the Netherlands

## Abstract

*The paper discusses the use and role of GIS-ILWIS software in determining conceptually the carrying capacity of the Upper Komering Sub Watershed for rice and coffee cultivation.*

*The carrying capacity was viewed and determined by the attractiveness of the land for cultivation. The attractiveness of the land was identified as a push factor of the farmer to expand and cultivate their land. Therefore, it determines the trend of the supply and demand for land ratio. The attractiveness was considered as the function of the supply of land and it was obtained by combining the physical suitability and socio-economic aspects i.e. the availability of physical infrastructure of the area and projected into different scenarios. The identification of the present and future demand of land was done by analysing the dynamic characteristics of the existing farming system i.e. farm size and change in farm size over time and population in the study area.*

*Accordingly, the GIS technology was employed to define and map out areas reflecting their carrying capacity. GIS has effectively handled the complexities of determining the carrying capacity based on user defined criteria, in this case, environmental or physical factors as well as socio-economic. The study yielded conclusive results that indeed GIS comes in handy to be employed in watershed management and development.*

## Introduction

This paper was written primarily to give a general descriptive analysis on carrying capacity in Upper Komering Sub Watershed, South Sumatera, Indonesia. The references used for this paper were mainly based on relevant reading materials on Resources Management and combined with those data collected and processed during the implementation of The Integrated Land and Watershed Management Information System (ILWIS) Project in Indonesia.

The paper will attempt to define carrying capacity by looking into the limiting environmental factors such as soil depth, nutrient contents, slope and climate. It affects the suitability of land for both coffee and rice production within the watershed. It will also attempt to put an added dimension by taking socio-economic factors such as land availability which has been determined as a function of actual land supply and demand as well as the accessibility of land, two other principal factors to be used to define carrying capacity of the watershed.

The authors of this paper would like to therefore to explore the possibilities of using GIS technology as supported by ILWIS software, as an effective tool has been employed as an effective tool with using criteria to define land suitability, availability and accessibility and further using these themes to determine the carrying capacity of the watershed through spatial data manipulation and analysis.

## II. Background and problems definition

Land ownership has been identified as one of problems in upper watershed development. It means that most of the farmers have limited access to the land, especially to the productive suitable and available land. In some regions, especially in Java, this phenomenon is much worse than the situation in Sumatera (Ramsay and Mulyadhi, 1984). Because of these factors population pressure on the land is so tremendous and pushed the landless or nearly landless farmers to expand their agricultural practices to very steep and mountainous areas which are not suitable for agricultural cultivation. Therefore, it is not surprising that some symptoms or negative impacts such as landslides, sedimentation, siltation of dams, drought and flood frequently occur in most of these regions. According to the Ramsay and Mulyadhi (1984), the erosion rate (2-3 mm/year) and natural soil formation capability (1 mm/year) in the study area are not comparable. These problems strongly indicate that the critical level of land resources exploitation or carrying capacity has been reached (Sumarwoto, 1985).

The upland agricultural land in upper watersheds is addressed mainly to hilly and mountainous areas that can be grouped and recommended generally into four types categories I to IV and restriction for all types of cultivation is given to the land with category IV (Ramsay and Mulyadhi,1984).

Based on this consideration the question about how far the available land can support and provide the basic needs of the people has arisen. Therefore the analysis of carrying capacity of watershed is very important aspect in order to give better insight in the situation of scarce land resources and in the formulation policy measures and other development interventions.

## III. General description of the study area

### Location and area extent

The largest part of the Upper Komering Watershed is located in Ogan Komering Ulu Regency, South Sumatera, Indonesia and lies approximately between lines of latitude 4 12 and 5 00 south and between line of longitude 103 35 and 104 27 east.

The study area covers 5 districts namely: (1) Banding Agung; (2) Muaradua; (3) Muaradua Kisam; (4) Pulau Beringin; and (5) Simpang. It is a typical humid tropical upper catchment with the annual rainfall range from 1780 mm to 3081 mm. The total area is approximately 400,000 hectare. Among the five districts only Banding Agung, Muaradua, and Simpang are well accessible by asphalt road.

### Land use and population in the study area

According to the Land Use and Vegetation Survey conducted by Surasana (1987), there are eight general vegetation-land use types: (1) rice/paddy field; (2) coffee plantation; (3) rubber plantation;(4) pepper plantation; (5) bushland;( 6) grassland; (7)savanna;and (8) forest.

The dominant land use-land cover is coffee (73% of the total area has been cultivated) as part of either shifting cultivation or permanent cultivation. The main food crop is rice and covers about 16 percent of the total area. It is usually cultivated in the valley bottoms while upland rice is grown in the hill, often incombination with young coffee plants.

Furthermore, based on the role of cultivated crops (main or additional crops) Suryana (1987) has classified the agricultural land into type of cultivation that represents the existence of farming systems in the study area which consists of coffee monoculture system (CMS) and coffee plus wetland rice (CWR).

According to Official Statistical Data (1984) the total population in this study area is 240.000 persons with average family size is 5 persons/household and growth rate 5 percent/year. This is relatively very high in comparison to the Indonesia annual population growth (2.1% per year). This figure indicates high substantial immigration rate especially in Simpang, Muaradua and Banding Agung District with high population growth rates. These districts also have a relatively dense road network. The population density is varies between 27 persons and 67 persons/km. The subsistence of the people is mainly dependent on the agricultural sectors (89% of the population).

## IV. Conceptual framework

### Carrying capacity defined

The term carrying capacity has been defined in different manner as viewed by various authors. According to Fearnside (1985) the human carrying capacity is the number of people that can be supported indefinitely in an area at a given standard of living, without environmental degradation taking place, given appropriate assumption concerning technology and consumptive habits. The foregoing definition of carrying capacity could be taken to include the concept of conservation, sustainability and productivity of the land to support the consumption needs and/or consumption patterns of a given population on an area over period of time. In other words, it implies the optimum and inherent capability of the land to produce various outputs to support and feed population within permissible level.

Geographers utilize carrying capacity as the maximum theoretical population that can be supported, given a resources base (Ricci, 1976). The approach employed in this case mostly consists of calculating the available agricultural land undergoing cultivation. Determination of the maximum population that an area can support is done by assuming a minimum acreage necessary to ensure survival of producing farmer.

EPA (1974) defines the term as the level of human activity (including population dynamics and economic activity) which a region can sustain as acceptable quality of life in the long term.

Geerling and de Bie (1986) define carrying capacity as a level of equilibrium between the availability of an element in an ecosystem (e.g. pastures) and the level of exploitation of that element (e.g. by herbivores). Clearly this definition of carrying capacity is basically related to the physical environment capacity.

Based on this perception, it is therefore possible for human carrying capacity to be analyzed through different approaches. As such the balance or imbalance between actual supply and demand for land could be considered as a basic indication for human carrying capacity analysis.

On the other hand, carrying capacity as a concept is often applied to indicate and estimate the level of permitted exploitation. Thus, in a sense, carrying capacity conveys an idea of limitation on the use of something.

### Carrying capacity analysis: an overview

The conceptual framework for this paper largely evolves on carrying capacity as previously defined. For the purpose of determining therefore the carrying capacity of the watershed, the term carrying capacity is defined in 2 (two) aspects: as a direct function of the physical environment, i.e. physical environment capacity, and the other aspect as a related function of socio-economic factors such as land availability as indicated by actual land supply and demand together with accessibility factors affecting land such as nearness to existing infrastructures.

Thus in this particular case, the following has been defined as a pre-requisite in determining the carrying capacity of the watershed and the setting up of the optimum level of exploitation: the area concerned is the whole watershed in general, and the terrain mapping unit in particular; the exploited element is land itself; while the object of exploitation in this particular case is the growing of either coffee monoculture or coffee plus wetland rice.

Hence, it could be assumed that the suitability of the land for combined growing of coffee and wetland rice would serve as framework of determining the physical carrying capacity. Therefore, all factors affecting land suitability will be classified as limiting.

Determination of land suitability was carried out by using FAO's guidelines and criteria which was combined with Land Evaluation Computer System (LECS) at the same time the spatial mapping of suitable and non suitable areas can be mapped out and was undertaken with the use of a GIS software called ILWIS.

This paper will also attempt to look at another equally important aspect of determining carrying capacity — that of identifying limit of exploitation as determined by socio-economic factors. In this case, socio-economic factors affecting land availability and land accessibility. Land availability has been defined as the total agricultural land available for cultivation. As such, factors that would affect land availability is basically the actual supply and demand for land, with demand being determined as direct function of farm household population, size of farm holding and change of farm size over time. Land supply is largely determined by the total area of the whole watershed less the already occupied areas.

Looking at farm household population, it could be also viewed as a limiting factor both as it affect demand as well as its effect on land productivity. Obviously, the higher the population, the higher would be the demand for land. On the other hand, diminished land productivity is also tied up to high population. The more the number of people there are, the higher would be the demand for food. Consequently, agricultural practices or farming system become intensive in order to satisfy and meet population demand for food. As farming system becomes intensive and without proper compensation for additional inputs to the land, the soil becomes poor in terms of depletion of soil nutrient available.

With regards to demand for land, this will be taken as indicative factor of the consumption pattern of the population. The premise would be that increase in population will mean an increase in the demand for agricultural land.

With these lead as framework, the study will yield 2 results of determining carrying capacity, that is, from purely ecological point of view which will make use of the land suitability factors and from a socio-economic point of view which will focus on factors affecting land availability and land accessibility.

## V. Method and techniques used

1. For the carrying capacity analysis purposes the data which has been collected by the ILWIS project especially farming systems data were used. All data manipulation was done using GIS-ILWIS commercial software.

2. Various criteria were formulated based on approved guidelines and FAO guidelines to be able to carry out suitability, availability, accessibility and attractiveness determination.

3. The attractiveness rating of the land has been done through procedures as stated as follows:

(a) Define land Suitability map for coffee

The land suitability map for coffee was derived by combining or overlaying two different maps i.e. the potential erosion under coffee and the ecological suitability for coffee map. By overlaying these maps five classes of suitability rating i.e. very suitable, suitable, moderately suitable, low suitable and non suitable were resulted.

(b) Define land accessibility map

The land accessibility map was derived by assigning classes of distance i.e. 0—2 km; 2—4 km; 4—5 km and >5km to the road map. The increasing distance per class represents the degree of accessibility.

(c) Define land availability map

The land availability map was derived from the landuse and landcover map through subtraction of the agriculture land (i.e. coffee or rice area) per terrain mapping unit by areas which have been occupied. By doing so, five classes of land availability i.e. 10 — 30%; 30 —50%; 50 — 70%; 70 — 90%; 90 - 100% were derived.

(d) Define land attractiveness

The land attractiveness map depicting land which are suitable, accessible and still available was resulted by overlaying three different maps i.e. land suitability, land availability and land accessibility map. The land attractiveness is consisted of five classes i.e. highly attractive, attractive, moderately attractive, low attractive and non attractive.

## VI. Results and carrying capacity analysis

### Results

Using GIS technology, an attractiveness map containing area labelled as "attractive in physical and economical context" were clearly defined. In coming up with this particular map, 3 (three) factors were largely considered, which were: suitability, availability, and accessibility. Based on these factors, a piece of land was determined as an attractive land if this land in general is suitable, available and accessible for cultivation and for growing coffee in particular. All of these 3 (three) factors involved in defining the attractiveness of land were duly considered as a function of land supply.

### Erosion hazard and land suitability map

For the factor suitability, the main criteria used focused mainly around agro-ecological suitability, that is taking into account the agricultural as well as ecological aspects of the land. For ecological related parameters, due consideration was given to soil depth, steepness of slope, and erosion hazard within the area. By and large, the soil formation rate within the watershed has been characterized as 1 mm/year while the soil depletion or degradation rate as function of erosion hazard, soil erodibility and steepness has been pegged at 3 mm/year (Ramsay and Mulyadhi, 1984). Clearly from the standpoint of carrying capacity, the critical point has been over reached. Soil depth was therefore considered a major limiting factor within the watershed and would largely dictate and influence agricultural cultivation. Steepness of slope was also a major consideration in the determination of land suitability. The higher the slope, the greater is the erosion hazard and risk. Thus, areas found at highest or steepest slopes (> 50 %) were considered suitable only as conservation area (protected forest). In other words, all land with steeper slope are less suitable for annual crops and require a lot of improvement. Inversely, areas at lower slopes are said to have lower erosion hazard and risk. They are generally with deeper soil, thus having higher suitability rating.

The resulting erosion hazard map and individual land suitability assessment for growing coffee and wetland rice in the study area are presented in the Table 1, and Table 2.

**Table 1:** Erosion Hazard Class Under Coffee in Upper Komering Watershed

| No | Erosion Hazard Class (ton/ha/year) | Description | Area (Ha) |
|---|---|---|---|
| 1. | 2.5 | Very low | 173,616 |
| 2. | 5.0 | Low | 108,294 |
| 3. | 10.0 | Moderate | 76,890 |
| 4. | 20.0 | High | 43,683 |
| 5. | 50.0 | Very high | 5,355 |
| | | Total | 407,838 |

**Table 2:** Land suitability for growing coffee or wetland rice in Upper Komering Watershed

| No | Land Suitability Class | Description | Coffee Area(Ha) | Wetland Rice* Area(Ha) |
|---|---|---|---|---|
| 1. | S1 | Very Suitable | 84,428 | 4,450 |
| 2. | S2 | Suitable | 168,258 | 6,803 |
| 3. | S3 | Moderate Suitable | 61,017 | 5,820 |
| 4. | S4 | Low Suitable | 47,435 | 36,304 |
| 5. | NS | Non-suitable | 46,700 | 354,461 |
| | | Total | 407,838 | 407,838 |

*) Source Sutaatmadja, 1987.

## Land availability map

Another factors considered in the determination of attractiveness of land is land availability. Land availability has been defined as the actual agricultural area available for cultivation. It involves all land which are not being cultivated. In our exercise, land availability information was adopted from Surasana (1987) which classified landuse in study area according to the percentage of dominant vegetation/cover on each TMU which did not distinguish between land being cultivated (exploited) and land being neglected (abandoned). In this exercise the availability of land can be seen as a direct result of the supply of land.

**Table 3 :** Land availability for cultivation in Upper Komering Watershed

| No | Land availability Class (%) | Description | Area (ha) |
|---|---|---|---|
| 1. | 10 - 30 | Very low | 36,939 |
| 2. | 30 - 50 | Low | 60,931 |
| 3. | 50 - 70 | Moderate | 93,485 |
| 4. | 70 - 90 | High | 133,221 |
| 5. | > 90 | Very high | 83,262 |
| | | Total | 407,838 |

## Land accessibility map

Accessibility as (an important) criteria for land attractiveness was also considered. In our case, accessibility of land was determined in terms of its "proximity or nearness" to an existing road as well as other support infrastructure facilities such as market, postharvest facilities, town, warehouse and processing zones. Thus, the nearer an area to such facilities, the higher is its accessibility class or rating.

**Table 4:** Land Accessibility for Cultivation in Upper Komering Watershed

| No | Land accessibility Class (km) | Description | Area (Ha) |
|---|---|---|---|
| 1. | 0 - 2 | Very accessible | 52,672 |
| 2. | 2 - 4 | Accessible | 40,159 |
| 3. | 4 - 6 | Moderate | 150,610 |
| 4. | > 6 | Less acessible | 164,397 |
| | | Total | 407,838 |

## Land attractiveness map

Examining the 3 major criteria in defining the attractiveness of an area revealed that the 3 has taken both the environmental and socio-economic factors. Environmental consideration are accounted for in the land suitability criteria whereas the socio-economic factors are embedded within the land availability and land accessibility criteria.

Taking all three aspects as factors to define the carrying capacity of the whole study area, resulting map which now clearly delineates the attractiveness of the land within watershed. We can nowdistinguish 36 categories each classified as to the degree with which the carrying capacity has been reached. To simplify the matter of categorizing the 37 classes and to be able to come up with meaningful data, it has been decided than the following simplified classification will be used:

| | |
|---|---|
| Highly attractive (A1) | Category 1 to 8 |
| Attractive (A2) | Category 9 to 16 |
| Moderately attractive (A3) | Category 17 to 24 |
| Low attractive (A4) | Category 25 to 32 |
| Non/Less attractive (NA) | Category 33 to 37 |

**Table 5:** The attractiveness of land for coffee or Wetland Rice Cultivation in Upper Komering Watershed

| No | Attractiveness Class | Description | Coffee Area (Ha) | Wetland Rice Area (Ha) |
|---|---|---|---|---|
| 1. | A1 | Highly attractive | 165,107 | 2,270 |
| 2. | A2 | Attractive | 102,356 | 9,486 |
| 3. | A3 | Moderately attractive | 58,205 | 19,300 |
| 4. | A4 | Low attractive | 60,245 | 16,633 |
| 5. | NA | Non attractive | 21,924 | 360,151 |
| | | Total | 407,838 | 407,838 |

*) Modified from Sutaatmadja, 1987.

### Carrying capacity analysis

*Carrying capacity analysis under agro-ecological land suitability perspective*

As indicated in the framework, the first procedure in defining the carrying capacity of the study area was to look into the environmental limiting factors. In this case, these were: Soil as characterized by physical (such as soil depth) or chemical soil fertility (such as pH), soil erodibility, rain fall/precipitation and the slope as it affect the growing of coffee within the study area. Coffee as an agricultural crop has its own specific requirements which were largely tied up to the soil type and its characteristics, the water requirement and altitudinal limits. This is basically a matching process wherein the crop requirement, in this case, coffee had been matched with inherent capability of the area. For this particular case, there were 5 suitability classes; very suitable, suitable, moderately suitable, and non-suitable. These classification has been arrived at using the predefined ecological factors such as soil fertility, soil depth, soil erodibility, rainfall and slope. The resulting land suitability classes for growing coffee and rice are presented on Table 2.

By determining the suitability of the area, we are also in effect determining the carrying capacity of the area since in essence suitability is equivalent to the physical carrying capacity because it concerns the same limiting factors, then we can draw the observation based on the results of the suitability map that areas classified as very suitable, suitable, and moderately suitable are the very area whose cultivation is still within the carrying capacity limits. It means that environmental hazard resulting from the cultivation area is rather minimal in view of the fact that the suitability mapping and matching has taken into account the erosion hazard of the area.

Either coffee or rice can be suitable on the same terrain mapping unit (TMU). To solve these dilemma the resulting suitability has been recalculated by assigning weights. In this case, if suitability rating on certain TMU for both crop are identical, then the priority is given for wetland rice. If the suitability rating for wetland rice is lower than coffee but the cost for improvement (alleviation of constraint) is lower than coffee again the priority is given to wetland rice. This particular decision has been resorted to inview of the fact that rice is the staple food of the population and therefore meeting the needs for rice is a priority in itself. Apart from that, wetland rice grows only on certain TMU in very limited area.

In terms of total area coverage for coffee cultivation, it means some 313,703 hectares (S1+S2+S3; see Table 2) of the total watershed area or about 77 % can be either being cultivated or not being cultivated are still within carrying capacity limit. On the other hand, those area falling within the low suitability for coffee cultivation (47,435 hectares) could be considered as areas approaching their carrying capacity limit when these areas are cultivated and as such necessary intervention in terms of tillage practices or farming systems modification should be introduced. But it might be that none of these areas are cultivated. For areas classified as non-suitable for coffee cultivation (46 700 hectares), it could be said that these areas if cultivated, these areas have already reached their carrying capacity limits in as much as they are now classified non-suitable area for the cultivation of coffee. It means that no cultivation of coffee should be undertaken simply due to the fact that these areas are characterized to be of very poor soil fertility and the agro-ecological zone is not fitted for coffee cultivation.

### Carrying capacity as a function of land attractiveness

*For existing demands*

The adopted working premised, was that the attractiveness of the land availability is related closely to the supply and demand for land it self. In this case, the higher the attractiveness and the larger the area of this class of land, the higher the demand and supply for this type of land. At that point in time where demand for land is being met with still considerable balance, we say that the carrying capacity has not yet been reached. Consequently, at that point in time where the demand for land could no longer be met and where there is a negative balance in terms of the supply and demand of land, we say that the carrying capacity has already been surpassed.

Why indeed these bold assumption? Accordingly, the demand for land is reflective of the total agricultural population growth, the farming system involved as characterized by their farm sizes and the changes in the farm size over time. [ As an assumption, we can say that in scenario "A", at time X, the population is P with average farm size as F ha and with demand for land D and supply LP. At time X, the demand D is sufficiently met with still an overall positive balance of LP - D ]. If LP> D, then that demand for food of the population can be met, and such human activities related to agricultural production is being undertaken within the limitation of the carrying capacity of the region.

In the other hand, if LP<D called scenario "B", then there is a negative balance as far as land demand - land supply ratio isconcerned. Food demand of population is now greater than actual food production. At this point in time, we can say that human activities have now reached the critical limit of the carrying capacity.

Going one step further, we now define carrying capacity as a function not only of land availability but of land accessibility as well. In this respect, a total of 92,831 hectares of land within the study area has been classified as accessible depending largely on the degree of accessibility and another 315 007 hectares has been classified as less accessible. Please find Table 4.

Overlaying this map with the suitability map, and availability map resulting map reflected suitable areas for coffee growing which are still available. With refer to Table 5 there are 5 categories in terms of attractiveness of the land. It is then highly possible that an area is suitable for coffee but it is no longer available, accessible and vice versa. The limiting factors and corresponding weight assigned considered were suitability (45%), accessibility (40%) and availability of land (15%).

Relating then these categorization to carrying capacity determination, we now arrive at the conclusion that those areas categorized as "Highly attractive" and "Attractive", these are the areas within which human activities expressed either by coffee or wetland rice cultivation is still within the carrying capacity limits. There are about 267,463 hectares of this types of areas within the watershed or roughly 66 % of the total land area which is suitable, available, and accessible for coffee cultivation and only about 11,756 hectares (2.9 %) for wetland rice cultivation (see Table 6 and 7). On contrary, those areas categorized as "Moderately attractive" and "Low attractive" these are the areas in which human activities both in coffee and wetland rice cultivation is approaching their carrying capacity limit. This roughly about 118,450 hectares (29%) for coffee and 35,933 hectares (8.9%) for wetland rice cultivation. In this areas, interventions should be started both physical and socio-economic ones. Now for those areas classified as "Non attractive", of which there is about 21,924 hectares (5% of the total watershed) for coffee and 360,151 hectares (88% of the total watershed) for wetland rice cultivation, these are areas are those in which human activities as well as coffee and rice production have already reached and exceeded their apparent carrying capacity limits. As such, the restoration of these areas should be of foremost importance and conservation measures should be carried out immediatly. Related to the carrying capacity analysis the attractiveness (suitable, available, and accessible) of land as mentioned above has been matched with demand for land. Demand for land in this case denotes actual individual needs for area to be cultivated to support and meet the basic needs on food per household level. Based on this consideration the demand of land can be quantified as a direct function of the size of farm holding, household farm population and the change in size over period of time.

Based on present data gathered (see Table 6) the actual demand for land both for coffee and rice in the year 1986 is 100,493 hectares (see Table 7) while the most suitable, available and accessible land (A1 + A2) is 267,463 hectares (see also Table 8) as indicated in the land attractiveness map which now mean that the demand for land is being met or satisfied by present supply of land. This could now be translated into the observation that as far as the human activiYties within the area is concerned, this has been apparent that their farming activities is still within carrying capacity level.

Based on attractiveness, a total of 267,463 hectares of roughly about 66 % of the total study area is highly attractive (A1) and attractive (A2) for coffee cultivation. Now taking into account the present population figures especially number of farm household (43,223 HH) and the current demand for land which is 100,493 hectares, then it is clear that the demand for land could still be met with yet a positive balance of 86,605 hectares. Further the average farm size for each farm household is about 1.8 hectares for coffee monoculture and 2.5 hectares for coffee plus wetland rice (Suryana, 1987). This is basically scenario "A" as was mentioned in the previous paragraph.

Using then projection for variable factors to come up with scenario "B", we can say demand for land especially for rice cultivation projected to be 18,478 (see Table 8), then it can be said that land supply for rice could no longer meet such demand. In addition, there is now a negative balance (7,091 ha) in terms of land supply - land demand ratio. Thus, in the year 1988 clearly it could be stated that human activities for rice cultivation within the study areas would have reached its carrying capacity limits.

**Table 6:** Projected number of households in agriculture

| Year | No of Household in | |
|---|---|---|
| | Coffee Monoculture (HH) | Coffee + Wetland Rice (HH) |
| 1986 | 23,773 | 19,450 |
| 1988 | 25,270 | 20,675 |
| 1990 | 26,860 | 21,977 |
| 1995 | 31,290 | 25,601 |
| 2000 | 36,450 | 29,823 |
| 2005 | 42,462 | 34,741 |
| 2010 | 49,464 | 40,470 |

Source: Mardanus, 1987.

**Table** 7: projected and additional demand of land*)

| Year | Demand | | | |
|---|---|---|---|---|
| | Coffee | | Rice | |
| | Projected (Ha) | Add. (Ha) | Projected (Ha) | Add. (Ha) |
| 1986 | 82,015 | - | 18,478 | - |
| 1988 | 87,937 | 5,922 | 20,882 | 2,404 |
| 1990 | 97,579 | 9,642 | 23,515 | 2,633 |
| 1995 | 116,626 | 19,047 | 33,281 | 9,766 |
| 2000 | 144,310 | 27,684 | 44,375 | 11,094 |
| 2005 | 180,655 | 36,345 | 60,102 | 15,727 |
| 2010 | 229,691 | 49,036 | 80,535 | 20,433 |

Note: *) Source modified Ibid, 1987.

**) Additional demand in year (t + n )= Demand in year (t+n) - Demand in year (t)

***) Year t = 1986

**Table 8:** The Actual and Projected Supply of Land

| Year | Supply*) | | | | | |
|---|---|---|---|---|---|---|
| | Coffee | | | Rice | | |
| | Att.1 (Ha) | Att.2 (Ha) | Total (Ha) | Att.1 (Ha) | Att.2 (Ha) | Total (Ha) |
| 1986 | 165,107 | 102,356 | 267,463 | 2,270 | 9,486 | 11,756 |
| 1988 | 83,092 | 102,356 | 185,448 | - | - | - |
| 1990 | - | 97,511 | 97,511 | - | - | - |

*) Supply of land in year t + n = land in year t - demand of land in year t + n

year t = 1986

**Table 9 :** Carrying Capacity (CC) Analysis Based on Supply and Additional Demand of Land for Coffee and Wetland Rice Cultivation

| Year | Total Supply | | Demand | | Balance | | CC Defined | |
|---|---|---|---|---|---|---|---|---|
| | Coffee (Ha) | Rice (Ha) | Coffee (Ha) | Rice (Ha) | Coffee (Ha) | Rice (Ha) | Coffee | Rice |
| 1986 | 267,463 | 11,756 | 82,015 | 18,847 | 185,448 | (7,091) | A | B |
| 1988 | 185,448 | - | 87,937 | - | 97,511 | - | A | B |
| 1990 | 97,511 | - | 97,579 | - | ( 68) | - | B | B |

Note: (....) = negative balance

A = Cultivation is still within CC

B = Cultivation has been surpassed CC

*For future demands*

Table 9 shows that in general the existing farming systems namely Coffee Monoculture and Coffee plus Wetland Rice system in study area would have surpassed the carrying capacity of the region untill the year 1990. It means that the agricultural land for coffee cultivation in this region will be very scarce by the year 2000. In other words, all agricultural land whereas categorized as available, suitable, and accessible land for coffee cultivation would have been occupied. But it is different for rice cultivation. Because in the year 1986 the carrying capacity of the region for rice cultivation will has been reached and had already negative balance. In other words, at that time then will be already a shortage of attractive land of 7,091 ha.

Our observation based on the foregoing discussion was that, indeed, that one could look at the determination of the carrying capacity in two perspectives: looking purely at the physical ecological frame work, and the other one relating carrying capacity to socio-economic factors. As we expected, there were two entirely different results when such assumptions was applied theoretically in determining the carrying capacity within the Upper Komering Sub-Watershed. However, we might as well add that what we did was largely subjective in the sense that we tried to come up with our own criteria to define the different categorization scheme. Nevertheless, we believed that in choosing the criteria we used, we have based them on measurable parameters.

## VII. Conclusion and recommendations

Based on the results of the foregoing, we would like to conclude and/or recommend the following:

A. On the use of GIS technology for carrying capacity determination:

1. GIS is a very useful tool in determining carrying capacity of an area principally because of the system capability to handle spatial and non spatial data which makes analysis more easier.

2. GIS capability to handle large volume of data sets and its inherent capability to compare mapped data in a one to one basis make it possible for the planner to come up with new data sets, the criteria of which are user defined. Thus, the use of GIS in determining carrying capacity of particular land could be effectively carried out.

3. GIS enables planners to come up with meaningful and easily understandable outputs — in the form of maps-which clearly define different area of interest that are very useful in planning and decision making.

B. On the determination of carrying capacity

1. Carrying capacity should be looked at not only in terms of the traditional way — that of determining carrying capacity of an area from a purely ecological point of view but also from a socio economic perspective such as land availability and land accessibility.

2. The combination of both ecological as well as socio-economic factors as a basis in carrying capacity determination should be employed, especially in areas where land for agricultural practices is limiting and population growth is relatively high.

3. Government agencies in developing countries, especially those who are involved in transmigration programs and policy formulation should take serious consideration on both ecological and socio-economic factors that affect carrying capacity of targeted areas for settlement purposes.

Finally, we recommed that GIS should now be integrated in programs/projects which are focused on determination of carrying capacity of any area. Furthermore, the integration of socio-economic factors in determing carrying capacity should be throughly looked into as it may have great impact on the carrying capacity limits of an area.

## Acknowledgements

Authors would like to thank: Prof. Molenaar and Prof. Meijerink for opportunity and advice; The Institute for Aerospace Survey and Earth Scientist (ITC) who allowed the use of Sumatera datasets; Dr.Valenzuela, Dr. Drummond, Dr. de Bie, Dr. Ketner and Mr. Brown Mr. de Meijere for their constructive discussion and suggestions; finally collegaues at Central of Geographic Information (CGI) and the Departement of Land Surveying and Remote Sensing, LUW especially Mr. Droesen, Mr. Buiten and Mr. Loedeman for encouragement and inspirations.

## References

Fearnside, P.M. 1985. A Stochastic Model for Estimating Human Carrying Capacity in Brazil's Transamazon Highway Colonization Area. *Human Ecology, Vol.3 No.3.* 1985. Plenum Publishing Corp.

Geerling, C. and S. de Bie. 1986. The Concept of Carrying Capacity and Landuse. *Netherlands Journal of Agriculture Science*, 34 334-347.

Mardanus, B. 1987. The Analysis of The Development of Farming Systems as an Elements of a Land Use Model for Planning Excercises in Upper Part, Komering Sub Watershed. MSc Thesis.ITC, Enschede. The Netherlands.

Sumarwoto, O. 1982. Ekologi, Linkungan Hidup dan Pembangunan. Penerbit Djambatan. Jakarta.

Suryana, N. 1987. Farming Systems and Their Distribution in The Upper Komering Watershed, South Sumatera, Indonesia. MSc Thesis. ITC, Enschede, The Netherlands.

Surasana, S.E. 1987. Agro-ecological zones of The Upper Part of Komering River Basin South Sumatera, Indonesia. MSc Thesis. ITC Enschede, The Netherlands.

Ricci, P.F. 1978. Policy Analysis Through Carrying Capacity. *Journal of Environmental Management.* 1978.6,85 97.

Ramsay,D.H. and D. Mulyadhi.1984. Report on Watershed Management in Indonesia. Excecutive Summary. MOA, MOF and Ministry Home Affairs in Cooperation with USAID. Jakarta, 1984.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Setting global and regional priorities for coastal temperate rain forest conservation

**Andrew Mitchell**

Conservation International
1015 18th Street, N.W., Suite 1000
Washington, D.C. USA 20036
Tel: (202) 429-5660
Fax: (202) 887-5188

## Abstract

*Coastal temperate rain forests are an uncommon major world biome restricted to the coastal margins of western North America, New Zealand, Tasmania, Chile, northwest Europe, and the Black Sea coast of Turkey and the former Soviet republic of Georgia. These relatively scarce forests are among the most productive in the world and are experiencing increasing pressures for resource utilization. However, information on the global or regional status of temperate rain forests is sparse. Identifying priority watersheds within each temperate rain forest zone, based on current levels of development and protection, will allow land managers to focus limited resources on those areas best suited for management as intact watersheds, forest management and development, or ecological restoration. This paper describes the development of a prototype GIS database for one region, the north Pacific coast of North America, with the purpose of establishing regional priorities for temperate rain forest conservation. The methodology used in identifying priority watersheds is discussed, as are issues involved in the development of a regional database which encompasses numerous jurisdictions. Future application of this prototype to other coastal temperate rain forest regions in an effort to establish a more accurate global information base for this biome is also discussed.*

## Introduction

*Coastal temperate rain forests.* Temperate rain forests are an uncommon major world biome restricted to the coastal margins of western North America, New Zealand, Tasmania, Chile, northwest Europe and the Black Sea coast of Turkey and the former Soviet republic of Georgia. Temperate rain forests harbor the greatest species diversity of any temperate ecosystems and are a priority for conservation (Harris 1984). Individual trees in these forests are long lived and tend to be the largest species in their genera (Franklin and Dyrness 1984). Coastal temperate rain forests are the base for regional timber economies and have great potential for long-term sustainable production. These forests harbor complex coastal watershed ecosystems that provide fish, wildlife and plant habitat (Eck 1984).

Although these ecosystems have been commercially exploited throughout this century, relatively little is known about the ecological condition of this uncommon biome, or its protection status globally. Recognizing the need to increase the level of knowledge about this endangered biome, Conservation International (CI), located in Washington, D.C., in conjunction with Ecotrust, an affiliate located in Portland, Oregon, has undertaken a project to map and analyze the status of the world's coastal temperate rain forests.

### Definitions

For this project, we are using a definition of coastal temperate rain forest developed by Weigand (1990). As stated by Weigand, "three features are common to all CTRFs [coastal temperate rain forests]: proximity to oceans, presence of high mountains (western Ireland being the sole exception), and, as a result of the interaction between the two, high amounts of rainfall." (Weigand 1990). This definition relies on the presence of vegetation, at least originally in a forested condition, lying within the temperate latitudes and a critical minimum annual rainfall amount. Weigand adopts 2000mm (80 inches) as this critical amount, with at least some rainfall occurring during the summer months (i.e., absence of an extended summer drought period). Alaback (1990) has identified presence of summer rainfall, "frequent disturbance by wind, and the lack of fire as an important factor in forest dynamics" as additional identifying features of temperate rain forests.

Within this definition Alaback (1990) has identified four major sub groups:

- Boreal temperate rain forest. Even annual rainfall distribution, cool summer temperatures (15 C), and considerable snowfall.
- Perhumid temperate rain forest. Even annual rainfall distribution, cool summer temperatures (15 C), and a minimum of snowfall.
- Seasonal temperate rain forest. Reduced summer rainfall, cool summer temperatures (15 C), and a minimum of snowfall.
- Perhumid subtropical temperate rain forest. Warm summer temperatures (25 C) and virtually no snowfall at sea level.

Using this definition, Weigand (1990) has identified five major coastal temperate rain forest regions:

- Northeastern Europe, including the western coasts of Ireland, Scotland and Norway, and the southern coast of Iceland.
- Eastern Black Sea, including the northeastern coast of Turkey and the southern coast of the former Soviet republic of Georgia.
- Tasman Sea, including the western coasts of Tasmania and South Island of New Zealand.
- Andean Austral, including the western coast of Chile south of Puerto Montt and extending inland into Argentina.
- Pacific Northwest, including the western coast of North America from northern California through the Gulf of Alaska.

## Ecosystem approach

In defining coastal temperate rain forests, Weigand and Alaback (Weigand 1990) emphasize ecological characteristics rather than species composition or forest structure. This approach allows for recognition of forests which are subject to similar environmental pressures, although occurring in widely separated areas of the world.

Mapping of global distributions of biome or ecosystem types in a theoretical manner has been initiated by Bailey and Hogg (1986), while at the continental scale ecoregion mapping has been undertaken (Bailey 1980, Omernik 1986) but the analysis of the global or regional status of specific biomes in relation to their management and conservation has not occurred. In addition, some sub-regional databases have been built, for example for National Forests in western Oregon and Washington (Morrison et al 1991). Within most coastal temperate rain forest countries outside of the Pacific Northwest region, research has focused on mapping of particular species rather than specific ecosystem types.

The purpose of this current research is to establish the distribution, development and protection status of the coastal temperate rain forest biome worldwide. The perspective provided by this project will allow conservationists to set priorities for temperate rain forest conservation and management both globally and regionally.

## An initial global overview

In an effort to establish baseline information on the original (pre-human settlement) distribution and extent of the coastal temperate rain forest (CTRF) biome, an initial GIS-based global overview was undertaken based on research performed by Weigand (1990).

### Methods

Based on his definitions, Weigand (1990) developed estimates of the original extent of coastal temperate rain forests worldwide using dot grid overlays and visual estimates of existing map manuscripts. For the most part, Weigand's estimates relied on precipitation, vegetation and forest maps at a variety of scales (ranging from 1:1,000,000 to 1:36,000,000) and map projections. The exceptions to this were the United States, where 1:100,000 scale maps were used, and Canada and Tasmania, where estimates from the British Columbia Ministry of Forests and the Working Group on Rain Forest Conservation were used, respectively. Weigand synthesized data from a number of sources and summarized this information by delineating the location and extent of coastal temperate rain forests on a series of small scale maps (ranging in scale from 1:2,000,000 to 1:36,000,000), except for the United States, Canada, Tasmania and New Zealand.

Given the limited scope and time constraints of the GIS-based global overview, and to maintain some consistency in scale, the small scale maps on which Weigand delineated areas of CTRF were digitized. For Alaska, British Columbia, Washington, Oregon and California, a 1:3,300,000 (approx.) map manuscript previously produced by Ecotrust/Conservation International was digitized. For Tasmania a map of forest types from the *Atlas of Tasmania* was digitized. For New Zealand, a map of historic extent of indigenous forests was digitized.

In many cases, the original map projection for these small scale maps was unknown, thus once digitized, the spatial accuracy was poor. To help improve the accuracy, maps of average annual precipitation for the CTRF countries were also digitized and overlaid with the previously digitized data. Most of these precipitation maps were from a series of climatic atlases published by UNESCO, at a scale of 1:5,000,000. Although the accuracy of this data may be questionable, the original map projection is known, and the projection and data is consistent within a region (e.g., Europe, North and South America). For Tasmania and New Zealand, precipitation maps produced in these countries, at scales of 1:10,000,000 and 1:2,000,000, respectively, were digitized. In all cases, only areas having average annual precipitation greater than or equal to 2000mm (based on Weigand's definition) were digitized.

The project utilized ESRI's PC ARC/INFO software running on a Hewlett-Packard Vectra 486 PC. Base map information (coastlines, water bodies, major rivers, and country boundaries) was obtained in digital format from ESRI's PC World Database, originally captured at a scale of 1:3,000,000.

#### Results

The final mapped areas of CTRF were based on what was deemed to be the most accurate fit between the three main sources of data: the areas of CTRF digitized from the small scale maps delineated by Weigand plus the source maps for the Pacific Northwest, Tasmania and New Zealand; the digitized average annual precipitation maps (from UNESCO, Tasmania, and New Zealand); and the digital base map information (coastlines and country boundaries) from the PC World Database.

Total estimated original extent of CTRF for each country, in hectares, was calculated in PC ARC/INFO using the digital maps. The total global original extent was calculated to be 41,177,500, as compared to Weigand's estimate of 30,954,184. There are several reasons for this discrepancy: the alternative methods used (dot grid overlay or visual estimate versus calculations performed by the GIS using digital maps); and additional or alternative data sources used in the two estimates.

While useful in producing a coarse baseline estimate of the original extent of the coastal temperate rain forest biome, the global overview underscored the need to develop more detailed analyses on a regional basis, including current distribution, development status and protection status of remaining coastal temperate rain forests.

### Regional priority watersheds analysis

The specific goal of the current analysis is to identify watersheds which are a priority for conservation within each temperate rain forest region based on current levels of human-based development (such as agriculture or logging activities) and protection status. This initial assessment will allow land managers to determine which areas are best suited for management as intact watersheds, which are best for forest management and development, and which may be candidates for ecological restoration. This landscape level approach will help determine the most effective use of limited conservation resources. The ultimate goal is to identify areas which can benefit from an ecosystem-based approach to economic development, thus providing a sustainable economic base for communities within the coastal temperate rain forest region.

To identify priority watersheds at the regional level, it is necessary to answer a number of specific questions for each coastal temperate rain forest region:

- What was the original distribution of coastal temperate rain forests?
- What is the current distribution of coastal temperate rain forests?
- Which primary watersheds does the coastal temperate rain forest zone fall within? Where are these watersheds and what is their size?
- What is the land cover composition of each of these watersheds?
- Which watersheds are developed, which are partially developed and which are undeveloped?
- Of the partially developed and undeveloped watersheds, which are fully protected, partially protected and unprotected?

Moore (1991) has attempted to answer similar questions for coastal British Columbia using a non-GIS based approach. Moore (1991) sites evidence that "an entire watershed is the most appropriate conservation unit in the coastal temperate forest". Reviewing primary watersheds larger than 5000 ha., Moore (1991) identified those which are developed (two percent or more of the total watershed area disturbed by logging, highways or other industrial development, or more than 250 ha. disturbed for watersheds greater than 10,000 ha.); those which are modified (less than two percent developed or less than 250 ha. for watersheds greater than 10,000 ha.); and those which are pristine (virtually no development with any logging less than five ha.). Moore (1991) also calculated the percentage of each watershed under protected status in a park or reserve. Similarly, in his global overview, Weigand (1990) estimated on a country-by-country basis the amount and percentage of coastal temperate rain forest in each of four classes: logged protected; unlogged protected; logged unprotected; unlogged unprotected.

Recent research (Hansen et al 1991) indicates that natural forests at a series of successional stages may provide more habitat benefits than managed forests. The focus of this project is to identify a broad coastal temperate rain forest ecosystem type which may in fact contain a variety of habitat types, and to distinguish between managed and natural forest in each primary watershed within the region. The amount of protected and unprotected natural forest within each watershed can then be determined.

## Pacific northwest regional database

As a prototype of this approach, a geographic database will be developed for the Pacific Northwest coastal temperate rain forest region, from Northern California through the Gulf of Alaska. Our knowledge base for this region is the most complete, and it is estimated that close to 40% of remaining coastal temperate rain forests occur in this region (Weigand, 1990).

### Methodology

Initially, primary watershed boundaries will be mapped, and watersheds greater than 5000 ha. in extent will be identified. Protected areas will also be mapped and classified based on level of protection provided (for example, National Parks in the United States provide a higher level of protection than do National Forests). Information on developed status of watersheds will be obtained using several alternative methods. Existing research, where applicable, (Moore, 1991; Weigand, 1990) will be included as tabular information and related to watershed polygons. In other cases, additional interviews with land managers will be conducted to determine developed status, and the resultant information included as tabular data. This tabular data will be corroborated through use of existing digital spatial data sets of land cover and forest class, which can provide surrogate measures of developed status (e.g., harvest date, crown closure, etc.), and through use of remotely sensed data such as NOAA AVHRR data and Landsat MSS data. Finally, through a series of map overlays, watersheds will be classified and mapped as developed, partially developed or undeveloped and by percent protected within several protection level classes.

### Database Development

The initial step is the development of the map database. In the interests of developing the regional overview quickly and minimizing the amount of spatial data that will have to be automated and managed, the project will utilize data from source maps at scales of 1:250,000 to 1:1,000,000.

Following a project database design phase, two pilot projects, currently in progress, one in the Willapa Bay and Grays Harbor watersheds of Washington and one along the north coast of British Columbia, will help evaluate the database design and provide valuable experience in dealing with data automation and conversion issues specific to the data available in this region. This will be followed by full scale data automation for the Pacific Northwest regional database, and subsequent spatial analyses and generation of final products. The prototype database is being developed using ARC/INFO GIS on a Sun Microsystems SPARCstation 2.

The emphasis is to acquire as much existing digital data as possible. Some data available only at larger scales will have to be generalized prior to inclusion in the database. Data not currently available in digital format will be digitized directly from source maps. Where existing digital data or source maps are unavailable, alternative methods, such as image classification of remotely sensed data, will have to be used.

There are several major themes of information included in the database. These are listed below, along with specific layers.

- Hydrography
    - Rivers
    - Lakes
    - Coastline
    - Watershed Boundaries
- Infrastructure
    - Transportation (Major roads, railroads and ferry routes)
    - Utilities (Major pipelines and transmission lines)
- Surficial
    - Elevation
    - Slope
    - Aspect
    - Bathymetry
- Administrative
    - Settlements
    - County Boundaries
    - State/Provincial Boundaries
    - Census Tract Boundaries
    - Land Status/Ownership
- Environmental
    - Land Cover
    - Forest Class

Average Annual Precipitation
Average Summer Precipitation
Average July Temperature
Ecozones

**Issues and Limitations**

Two major issues in development of the prototype regional database have been the lack of reliable spatial data for some specific data layers in portions of the region, and the integration of existing digital data sets from various sources. Current digital data at regional scales on land cover, vegetation and forest class are unavailable for portions of the Pacific Northwest region. While a common problem in tropical rain forest regions, this has been somewhat unexpected for the temperate rain forest region. Efforts are currently underway by several private organizations and governmental agencies to develop this critical information, primarily using satellite imagery.

Integration of existing digital data sets has presented a number of issues which need to be resolved to create a useful database. These include processing data of varying scales and resolutions, differences in definition (e.g., criteria used to define ecoregions in the United States and biogeoclimatic zones in British Columbia), and varying classification schemes for data (e.g., forest class types on National Forest versus state lands). In many cases, for this regional analysis, generalization of spatial data and simplification of classes for attribute data has allowed integration of the various data sets.

This generalization, while necessary at the regional scale, still allows analysis consistent with the project objectives. However, this data is not applicable to site specific management of particular watersheds or development of management plans. The development of more detailed databases is required to support this type of analysis. As specific conservation priorities within the temperate rain forest region are identified, regional datasets can provide an initial framework and information base for development of the local database. As detailed datasets are developed at the local scale, they can be patched into the database and also be used to update and improve the regional database.

## Future regional databases

The regional database and analysis methods developed for the Pacific Northwest region can serve as a prototype for development of similar databases in other coastal temperate rain forest regions, thus leading to the identification of priority watersheds in these regions.

The approach to implementation of these databases will vary depending on the technical and political environment within the specific region. In Tasmania and New Zealand, for example, much digital data exists and core GIS databases have been developed. In addition, several governmental agencies and private organizations are actively engaged in rain forest research. Thus the prototype database could become an additional tool in this ongoing effort. In Chile and Argentina, on the other hand, while interest is high, little research specifically on coastal temperate rain forests has been conducted, and digital data sets are scarce. In this region, the use of AVHRR and other remotely sensed data will acquire greater significance. Here, the prototype database could provide a framework for data collection efforts and basic research, as well as eventual application as an important analysis tool.

It should be noted that there may be limitations to the regional approach specific to particular coastal temperate rain forest regions. In Chile, for example, most development of the forest has historically consisted of clearing of small plots (Lara, 1987). These will be difficult to identify and map at the regional scale. In any case, it is critical that future development of regional databases be closely coordinated with conservation organizations, universities and governmental organizations within the region, and that development of ecosystem-level databases, analyses and management plans be generated locally.

## Conclusion

Development of a regional database using data at scales of 1:250,000 to 1:1,000,000 can provide an overview of the status of the coastal temperate rainforest biome within the Pacific Northwest region. This regional overview can help land managers identify those watersheds which are a priority for conservation and which should be targeted for additional research. Recognizing that there is a fine line between overview and oversimplification, given the primary project objective of building a regional context for conservation, it is beneficial to produce such a landscape level overview.

The context provided by the regional overview can support development of local, ecosystem-level databases at scales of 1:24,000 to 1:250,000 to help develop long-term ecosystem management plans, perform ongoing ecosystem monitoring and perform project-specific analysis for specific priority watersheds. The prototype regional database may also have some applicability in other regions within this endangered and highly productive biome.

## Acknowledgements

This project has received generous support from Sun Microsystems, Environmental Systems Research Institute, Tektronix, and CalComp.

## References

Alaback, P.B. "Comparative ecology of temperate rain forests of the Americas along analogous climatic gradients." U.S. Forest Service, Pacific Northwest Research Station, Juneau, Alaska, 1990.

Bailey, R.G. *Description of the Ecoregions of the United States*. United States Department of Agriculture, Forest Service, Miscellaneous Publication 1391, 1980.

Bailey, R.G. and H.C. Hogg. "A World Ecoregions Map for Resource Reporting", *Environmental Conservation*, Vol. 13, No. 3, Autumn, 1986.

Eck, K.C. "Forest characteristics and associated deer habitat values, Prince William Sound Islands," in W.R. Meehan et al., eds. *Fish and wildlife relationships in old-growth forests*, American Institute of Fishery Research Biologists, Morehead City, North Carolina, 1984.

Franklin, J.F. and C.T. Dyrness. *Natural Vegetation of Oregon and Washington*, Oregon State University Press, 1984.

Hansen, A.J., et al. "Conserving Biodiversity in Managed Forests", *BioScience*, Vol. 41, No. 6, June, 1991.

Harris, L.D. *The Fragmented Forest: Island Biogeography Theory and the Preservation of Biotic Diversity*, University of Chicago Press, 1984.

Lara, A.A. and G.M. Cruz. Evaluacion del potencial de pastoreo del area de uso agropecuario de la XII region, Magallanes y de la antartica Chilena. Investigacion Tecnologia Agropecuaria XII Region, Instituto de Investigaciones Agropecuarias, Santiago, Chile, 1987.

Moore, K. *An Inventory of Watersheds in the Coastal Temperate Forests of British Columbia*, Earthlife Canada Foundation and Ecotrust/Conservation International, 1991.

Morrison, P.H., et al. *Ancient Forests in the Pacific Northwest*, The Wilderness Society, 1991.

Omernik, J.M. *Ecoregions of the United States*, Corvallis Environmental Research Laboratory, U.S. Environmental Protection Agency, 1986.

Weigand, J.F. "Coastal Temperate Rainforests Definition and Global Distribution with Particular Emphasis on North America." A report prepared for Conservation International, Portland, Oregon, 1990.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Evaluating alternative agricultural land management policies from a water quality perspective: An automated land information systems approach

**Lucy A. Buchan**
**Peter G. Thum**
**Bernard J. Niemann, Jr.**
**Stephen J. Ventura**

Land Information & Computer Graphics Facility
School of Natural Resources
College of Agricultural & Life Sciences
University of Wisconsin - Madison
Madison, WI 53706

## Abstract

*The integration of geographic and land information system (GIS/LIS) techniques and computer modelling is increasingly being utilized to support rural resource management and policy analysis. GIS/LIS data handling capabilities can enhance tabular models by incorporating dimensions of space and time. Spatial analyses can augment the understanding of relationships between landuse and environmental quality. They have been proven to be particularly helpful in implementing and assessing the impact of program mandates such as the 1985 Food Security Act (FSA) and 1990 Food, Agriculture, Conservation, and Trade Act (FACTA). These mandates link environmental quality standards to agricultural practices and U.S. Department of Agriculture (USDA) Commodity Program benefits. This paper presents the results of a study designed to evaluate the contribution of conservation treatment and agricultural forests in maintaining environmental quality. The study was conducted in the Garfoot Creek drainage area of the Black Earth Creek Watershed (Dane County, Wisconsin).*

*The study explored the usefulness of a creating a multipurpose LIS and linking it to a water quality assessment model for evaluating existing and pending land management policies. In order to develop the model land management scenarios used in these evaluations, numerous natural resource layers and land management practices were integrated into a digital database. A land cover data layer was developed using experimental orthophotography. Forest cover was analyzed to determine its spatial and temporal distribution. Information about agricultural practices at the farm field level was obtained from the Dane County Land Conservation Department's (DCLCD) farm cooperator database. This data, combined with information on soils, slope, hydrology and landcover, was used as input to the Wisconsin Nonpoint Source Model (WIN). The model, used by the Wisconsin Department of Natural Resources (DNR), estimates soil erosion at the individual farm field level and calculates sediment delivery based on surface water flow between individual farm fields. Agricultural use scenarios were developed, modeled, and evaluated to determine the environmental outcomes associated with each scenario. Reductions in agricultural nonpoint pollution due to changes in forest cover extent versus traditional conservation treatment are compared.*

## Synopsis of results

The results of our investigations indicate that traditional conservation treatment practices can significantly reduce soil erosion and sediment delivery rates on highly erodible lands. However, the introduction of forest uses provides even more significant reductions when it replaces highly erosive agricultural practices. Also, if all highly erodible cropped fields (where actual soil loss exceeds tolerable conditions) are brought within tolerable limits using traditional conservation practices, reductions in soil erosion and sediment delivery would be comparable to changing cropped use to forest use.

The study also found that despite more land in crop production, soil erosion and sediment delivery rates in the 1930's were similar to current conditions with conservation treatment. This appears to be primarily due to the less intensive crop rotations used in the early part of the century. The study suggests that the WIN model is a more reliable and valid measure of water quality impact than the Universal Soil Loss Equation (USLE) since it allows policymakers to estimate sediment delivery to receiving waters. GIS/LIS data manipulation and analysis can greatly facilitate the evaluation of land management options using tabular water quality models such as WIN.

## Introduction

Until recently, policies that influenced rural land use encouraged traditional agricultural production and timbering, even on sites where the soil type and slope made practices detrimental to the environment. Little or no consideration was shown for other concerns such as watershed protection, biological diversity, and wildlife and fish habitat. Since the 1970's, there has been growing interest in identifying the external impacts of human resource use by better understanding physical landscape processes and interactions. This knowledge can provide the foundation for policies that require rural landowners to manage their land not only to meet production and economic goals, but to also protect environmental quality. The research goal of this study was to evaluate the utility of using GIS/LIS technology to help quantify the benefits of sound rural resource management as called for in the conservation provisions of the 1985 FSA and 1990 FACTA, establish a sound basis for evaluating long-term tradeoffs in various land management options, and to evaluate policies that advocate the investment of public funds in rural resource management. GIS/LIS techniques combined with agricultural nonpoint pollution models provided a means of analyzing the impact of existing land uses and comparing them to alternative land use management scenarios. These predictive analyses offer resource planners and policy makers "hard" data to evaluate the impact of resource policy and build an awareness among rural landowners of their influence on environmental quality.

During this study, special emphasis was given to the role of agricultural forests in environmental quality. Apart from providing commercial products, scenic beauty and visual contrast in a landscape, agricultural forests offer habitat to wildlife and deter soil erosion and sediment delivery, thereby protecting surface water quality and associated fisheries. This study utilized water quality for brown trout as a measure of environmental quality since Garfoot Creek, the study area, is tributary to Black Earth Creek — known for its high quality trout fishery. By examining this measure, the relative contributions of land use for agricultural production or as forest cover, were compared. Research showed how these different land uses could reinforce one another and also quantified their relative contributions to the selected measure of environmental quality.

Water quality was measured as the amount of sediment delivered from cropped farm fields to surface waters. In order to best serve decision-making by resource planners, managers and policy makers, it is essential to provide information that relates sediment delivery to individual agricultural operators and their unit of management (i.e. farm field). Therefore, the WIN model is designed to utilize a database that includes information at the farm field level and associates each field with a farm operator. Output from the model identifies critical upland areas contributing sediment to receiving waters which permits more effective targeting of soil erosion and/or sediment controls. Running the model under various land management scenarios provides resource planners and policy makers with a flexible tool for deciding how to achieve environmental quality goals while still meeting the production needs of agricultural operators.

## Study area

Research was conducted in the Garfoot Creek subwatershed in the township of Cross Plains (Dane County, Wisconsin), located within the Black Earth Creek Watershed. The drainage, encompassing approximately six square miles, is a complex mixture of glacial river terraces, steep side slopes, and flat ridges. The original landscape was composed of mostly prairie, wetlands and forests. It is now intensively farmed on terraces and ridges while slopes are mostly second-growth mixed oak forests.

## Methods

### Database development

Establishment of a multipurpose LIS for this project required automation of many natural and cultural features as individual data coverages. Database development work began during an earlier phase of the research which evaluated wildlife habitat suitability as a measure of environmental quality. A descriptive model was developed using GIS/LIS to simulate expert knowledge

**Table 1: Data sources and processes used to create coverages**

| Coverage | Data Source | Automation Process |
|---|---|---|
| Sections | USGS 1:24000 DLGs | Digitized |
| Townships | USGS 1:24000 DLGs | Digitized and dissolved on Sections |
| Roads | USGS 1:24000 7.5' Quadrangle Sheets | Digitized |
| Soils | SCS 1:15840 Soil Survey | Soil sheets digitally scanned |
| Slopes (TIN) | USGS 1:12000 3.75' Orthophoto Quadrangle Sheets | Digitized elevations into point file. TIN software created polygon coverage. |
| Streams | USGS 1:24000 7.5' Quadrangle Sheets | Digitized |
| Wetlands | WDNR 1:24000 Digital Database | No further work required |
| Farm Fields & Tracts | ASCS 1:7920 Unrectified Aerial Photography | Digitized field & tract boundaries. Ascii file transfer of attribute data into digital coverage. |
| Landcover | 1. 1931 Land use economic inventory map<br>2. USGS 1:12000 3.75' Orthophoto Quadrangle Sheets | 1. Digitized<br>2. Manual interpretation and entry into Field and Tract Coverage. |

on habitat appropriate for the re-introduction of wild turkeys (Serra et al., 1990). Additional coverages were required for the water quality study including reclassified landcover, farm field and tract boundaries, hydrology, and associated attribute data describing vegetation type and agricultural management practices. A summary of data sources and processes used to create final project coverages is provided in Table 1.

*Farm field and tract (ownership) boundaries*

Farm field and tract boundaries, based on unrectified aerial photos maintained by the USDA Agricultural Stabilization and Conservation Service (ASCS), were digitized using pc ARC/INFO. These boundaries are delineated by both ASCS and DCLCD personnel to meet their respective federal and state mandated program planning and compliance monitoring requirements. Farm fields served as the unit of analysis for the WIN model and provided the link to crop management attributes maintained in the tabular DCLCD database, and other landscape characteristics (i.e. soils, slope). Landowners enrolled in any DNR forestry programs were identified by reviewing ASCS farm management records and DNR files. Enrollment in a program was included as an attribute in the database and served as a measure of future land use intent.

## Field management

The DCLCD 's tabular database contained detailed information about farming practices at the farm field level. Information is stored on a field basis and is linked to ownership by unique id's assigned to each farm operator. Pertinent crop management information was exported from the database as an ascii file and then linked to the study area field coverage based on the unique field id. Technically this is a simple process, however constant changes in tract and field ids by ASCS resulted in incomplete or old information being obtained from the DCLCD database. Therefore, one had to constantly cross-reference mapped field and tract ids with ASCS changes and DCLCD database updates. For a small area such as Garfoot Creek subwatershed, it was just just as easy to write field attributes to hard copy (printout) and manually code them into the field coverage. The type of attributes linked to the field coverage included crop type, planting rotation, plowing practices, and the tolerable soil loss assigned to each field based upon predominant soil characteristics.

*Soils*

The Dane County soil survey was automated prior to this study as part of a larger research project — the Dane County Land Records Project. Digital soil boundaries for the study area were clipped from a larger coverage and then overlaid with the farm tract and field coverage. This resulted in several soil types per field. Since WIN requires only one soil type designation per field, the predominant one based on area was assigned. Additional soil characteristics transferred to each field in the field coverage through the overlay included percent slope, soil erodibility (K), and a SCS defined slope-length parameter (LS). Soil characteristics available in the DCLCD database superceded those determined through the overlay because they were based on site field investigations of farmers participating in conservation programs.

### *Landcover*

Landcover was originally developed to support the wildlife habitat investigation. A vegetation map from 1832 (Finley, 1951), a 1931 land use inventory map drafted during an economic land inventory (Kaufer, 1976), and 1987 landcover interpreted from the 1:12,000 orthophotographs were automated. The three maps were digitized and transformed into a local Transverse Mercator projection. For standardization purposes, the landcover categories were aggregated to represent the simplest of the three classification schemes; the Finley map.

Landcover categories consisted of forest, wetland, and cropland. Oak opening and prairie cover were independently digitized based on more detailed survey notes written by Finley that were not included in his final map. In the 1931 and 1987 coverages, the category of abandoned field/pasture was added to illustrate spatial land use changes. The 1987 forest cover was further classified by interpretation of the 1:12000 orthophotos to reflect the forest types represented in the WDNR forestry programs, ('Woodland Tax', 'Managed Forest Law', and 'Forest Cropland'). The general forest category was expanded into either mixed oak, red cedar or pine.

Temporal changes in landcover were measured to assess spatial distribution of suitable wildlife habitat and predict its distribution in the future. Analyses revealed that forest cover shifted from 56% coverage in 1832 to 29% in 1931 and to 49% in 1987. Another ongoing phase of research (Heinzel, 1991), is utilizing automated methods to analyze patch sizes, corridors, forest edges, and edge to area ratios to further understand the implications of changes in land use and their impacts on ecological resources.

Specific needs of the WIN model required reclassification of this landcover to include additional cover categories such as urban, shrub, grassland, and pasture. Landcover was modified by manual interpretation of the 1:12000 orthophotos. A mylar plot depicting the subwatershed boundaries, roads, streams, and field and tract boundaries at 1:12,000 was overlaid on the orthophotos. Landcover types were classified, delineated on the plot and then added as features to the vegetation coverage. For the 1931 agricultural management scenario, that years' economic inventory was used.

### *Flow routing*

Sediment delivery calculations performed by the WIN model require surface water flow information. This information takes the form of user defined flow linkages between individual farm field polygons. Since water flows from a field into either another field or water body, the terminology used to describe this attribute is "outcell". The outcell and flow length across the field are included in the field and tract coverage.

Previous research for the study area had produced a digital point coverage in ARC/INFO format containing elevation values derived from USGS 7.5" quadrangle contours. These points were used create a triangulated irregular network (TIN), defining terrain slope and aspect characteristics. A Fortran algorithm was then developed and used to generate a separate coverage of lines based upon the aspect of each triangle in the TIN. The lines ran parallel to the aspect of each triangle, and when graphically displayed depicted the surface water flow direction across an area. To relate this flow routing information to study area farm fields, a graphic overlay of the field and tract coverage, aspect lines coverage, and stream coverage was created. The graphic facilitated interpretation of field flow linkages (Figure 1).

Flow linkages were assigned to each field through visual examination of the field/aspect line graphic and database coding. Direction of flow was estimated by judging the direction of the majority of the flow lines on a field. Large fields that exhibited complex topography required subdivision according to smaller topographical units. In this case a common farm field id was maintained for each subfield. A separate Fortran algorithm was written to try and automate the interpretation process, however, the program was still not bug free at the time of project documentation. Programming obstacles are attributable to the fact that it is much more difficult to handle surface water flow routing in a vector data structure than in a raster (grid cell) environment.

## Estimating sediment delivery using the WIN model

### *Model operation and inputs*

The WIN model estimates rates of sediment delivery from rural lands to surface waters. Streams can be divided into segments requiring special study. WIN contains options to estimate sediment delivery resulting from single or multiple rainfall events. Rates are associated with individual fields and are linked to ownership through farm field and tract attributes in the database. The model can be used interactively by resource managers to evaluate alternative land management practice scenarios. It can also be used to evaluate the impacts of various conservation treatment options. For example, a county resource manager may want to estimate the quantity and cost of practices needed to reduce in-stream sediment loads to target levels, as well as the time needed to design, install and certify these practices. If model results for a particular land management scenario are unacceptable (i.e. they do not meet established standards or policy), the land management plan can be quickly modified and the model run again to evaluate the effects of these changes. Alternatively, policy makers may wish to modify policy if required practices place excessive management or economic burdens on farm operators/owners. The effect of policy changes can also be quantified using WIN.

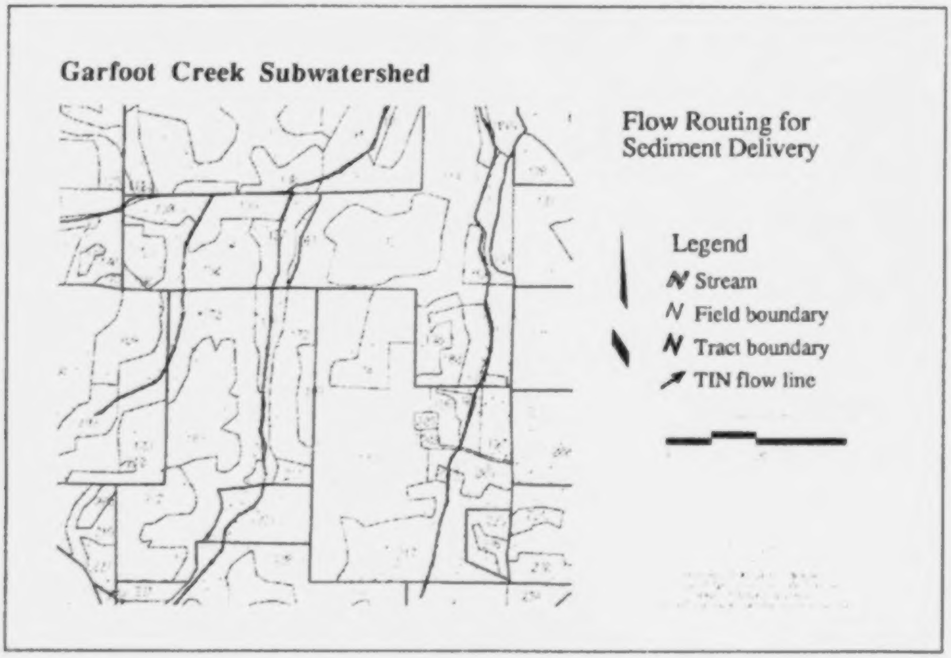

Garfoot Creek Subwatershed
Flow Routing for Sediment Delivery
Legend
Stream
Field boundary
Tract boundary
TIN flow line

**Figure 1**

Potential farm field soil loss due to sheet and rill erosion are calculated by the model using an adaptation of the SCS Universal Soil Loss Equation (USLE). The variables required by USLE are listed in Table 3. The USLE has traditionally been used by resource managers as a sole indicator of agricultural environmental quality. In order to use this tool as a measure of water quality, a number of assumptions have to be made about how much sediment will actually leave the field, its rate, flow path, and eventual destination. WIN is able to estimate sediment delivery rates and receiving water impacts by by taking into consideration many more landscape variables then the USLE.

**Table 3:**
**The universal soil loss equation (USLE)**

$A = R * K * LS * C * P$

| | |
|---|---|
| A | Predicted average annual soil loss in tons/acre/year |
| R | Rainfall factor |
| K | Soil erodibility factor |
| LS | Slope-length and slope-gradient factor |
| C | Cropping management factor |
| P | Erosion control practice factor |

WIN input requirements include information about soils, agricultural practices, landcover types, hydrology, topography, seasonal rainfall, and climatic factors effecting evapotranspiration rates. For our study area, rainfall and evapotranspiration data was available from the Wisconsin DNR. Other data was extracted from the coverages described earlier. Table 5 lists input parameters to the model, their data source, and a brief description of each parameter. Figure 2 provides a schematic summary of the WIN database compilation process.

Not all field attribute data compiled for this study was required input of the WIN model. Some attributes were included for graphic display and enhancing interpretation of the land use scenarios. A number of pcARC/INFO macros were written to extract the field attributes required by the model and sort them in proper input order. An intermediate ascii file was used to transfer the data from pcARC/INFO to the Rbase database management program. WIN uses the Rbase format for internal data manipulations.

In addition to field data, separate Rbase files were created to serve as lookup tables for the model to cross-reference crop rotations, soil attributes, watershed

**Database Compilation Process for the WIN Model**

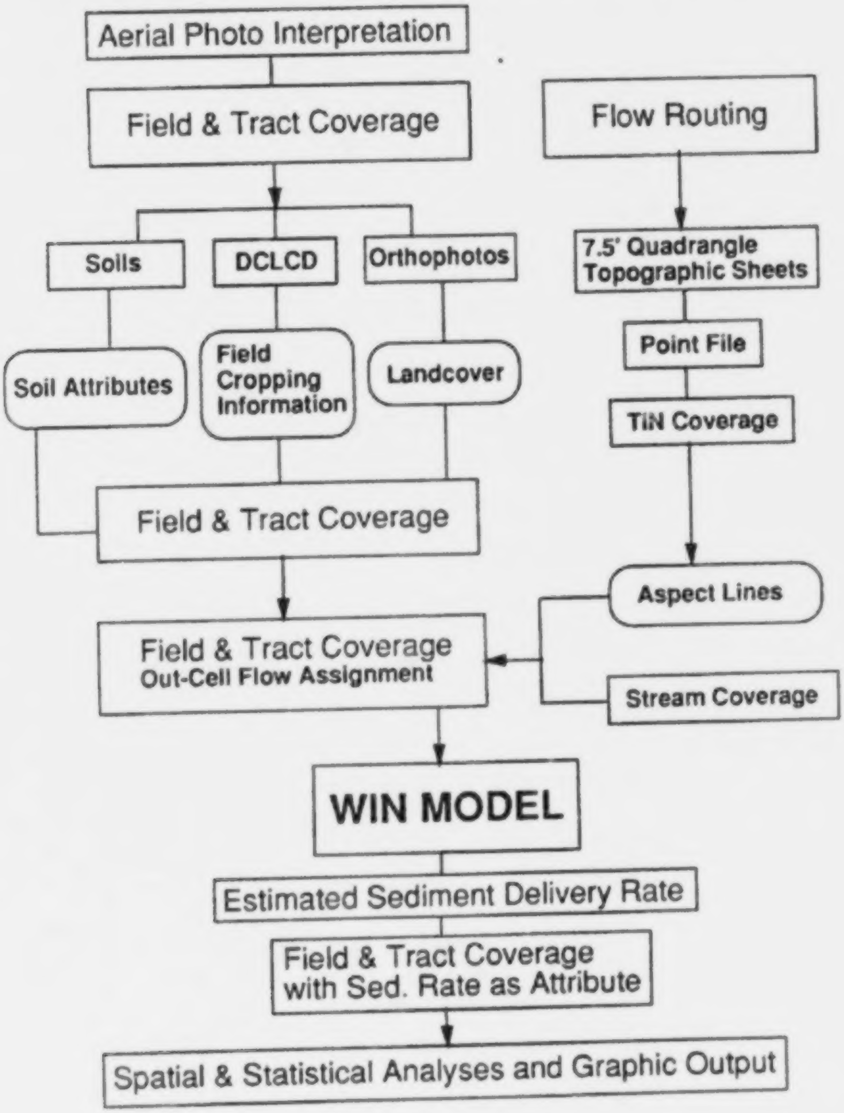

Aerial Photo Interpretation
Field & Tract Coverage
Flow Routing
Soils
DCLCD
Orthophotos
7.5' Quadrangle Topographic Sheets
Soil Attributes
Field Cropping Information
Landcover
Point File
TIN Coverage
Field & Tract Coverage
Aspect Lines
Field & Tract Coverage Out-Cell Flow Assignment
Stream Coverage
WIN MODEL
Estimated Sediment Delivery Rate
Field & Tract Coverage with Sed. Rate as Attribute
Spatial & Statistical Analyses and Graphic Output

Figure 2

**Table 5: Description and source of data input to WIN model**

| INPUT PARAMETER | DATA SOURCE | DESCRIPTION |
|---|---|---|
| Ssaid<br>Soil Symbol<br>Slope %<br>K,LS,C,P,R | Soil<br>Conservation Service | -Soil Code #<br>-Soil Type<br>-Percent grade of soil<br>-Parameters used in USLE pertaining to soil texture, slope,plowing and cropping mngmt., and avg. rainfall |
| Farm Operator<br>Tract #<br>Field #<br>Crop Rotation<br>Rotation Code<br>Plow Practice<br>Mgmt. Code<br>T | Dane County Land Conservation Office (DCLCO) | -Names and id's used in the Dane County Cooperator Tracking System<br>-Crop and planting pattern<br>-Code for crop rotation<br>-Up-and-down, contour, strip<br>-Conservtn. tillage practices<br>-Tolerable soil loss |
| Landcover | USGS Orthophoto 7.5' Quadrangle Sheets | -Landcover types identified from orthophotos ( Cropped fields, Pasture, grassland, forest, wetland, residential) |
| Flowlength<br>Celltype<br>Cellout | Fortran Flow-routing algorithm using:<br>1. Field & Tract,<br>2. Stream, and<br>3. TIN coverages | -Length of Field that sediment is likely to flow<br>-Whether the cell of concern is a field or a stream<br>-Whether the receiving cell is a field or a stream |
| Watershed Code<br>Subwatershed Code | Wisconsin Dept. of Natural Resources | -Code assigned to watershed.<br>-Code assigned to sub-watershed |
| Township Code<br>In-cell township<br>Out-cell township | Township Coverage from USGS 1:24000 DLGs | -Code assigned to township twice, once for in-cell, once for out-cell |
| Section Code<br>In-cell section<br>Out-cell section | Section Coverage from USGS 1:24000 DLGs | -Code assigned to PLSS section, once for in-cell, once for out-cell |
| USLE<br>Area<br>Parcel ID | Calculated in Database | -USLE = K * LS * C* R * P<br>-Area of parcels in Acres<br>-Unique Identifier # |

statistics, etc. Model runs for the study area took less than thirty seconds for each scenario, though data debugging usually added another hour or two to each scenario run. Model output included estimates of sediment delivery for each farm field and for non-cropped fields such as pasture and forest. WIN has default values for non-cropped uses. Model output also included total subwatershed sediment production for individual storm events in an average year. Rainfall duration, depth, and frequencies were statistically generated using DNR guidelines. Farm field sediment delivery results were linked back to the field coverage for graphic display and interpretation.

*Land use scenarios*

The Garfoot Creek subwatershed farm tract and field coverage was manipulated to represent five different agricultural land use scenarios. They reflected different acreages of forest and cropland in the subwatershed as well as various agricultural practices.

Agricultural Use Scenario 1. This scenario portrayed existing land use with in-place agricultural conservation practices. The scenario provides a reference point for comparison with other land use scenarios.

Land use scenario 2. This scenario reflects existing agricultural use with none of the soil conservation provisions mandated by the 1985 FSA and 1990 FACTA. Comparison with Scenario 1 quantifies the impact of these soil erosion conservation provisions.

Land use scenario 3. This alternative depicts the contemporary rural landscape where cropped land is allowed to revert to forest when actual soil loss (A) is greater than the tolerable soil loss (T).

Comparison with Scenarios 1 illustrates the impact of utilizing agricultural forest to prevent soil erosion and sediment delivery, relative to the implementation of agricultural conservation practices.

Land use Scenario 4. This alternative describes the contemporary rural landscape where cropped land with agricultural practices that experience actual soil loss exceeding tolerable conditions (A > T) are amended to yield A = T. Amendments are in the form of traditional conservation treatment instead of conversion to forest use. Comparison with Scenario 1 measures the effectiveness of existing conservation practices. Comparison with Scenario 3 highlights the difference between forest cover and conservation treatment for controlling soil erosion and sediment delivery.

Land use scenario 5. This scenario recreates the rural landscape as it appeared in 1931, using the landcover and agricultural practices typical of that period, but assuming current field and tract delineations. The purpose was to depict a situation similar to the Hubbard Brook Study (Bormann, et al, 1974; Likens et al, 1977) where an entire subwatershed was deforested to measure the amount and rate of soil erosion. This study uses computer and GIS techniques to conduct a similar experiment, though complete deforestation is not assumed. The 1931 economic land inventory map waschosen as a more realistic alternative; 1931 being a time where forest cover acreage in Wisconsin was at its lowest since European settlement.

The decision to use current field and tract boundaries for the 1931 landcover was made to provide clear graphic output for visual comparison. The 1931 landcover distribution was assigned to the current field and tract delineations by a majority-based rule. Whichever landcover category comprised the greatest acreage was assigned to the field. Such class-ification does alter the true landcover distribution for that time period, but it also does not accurately reflect surface water flow conditions. Subsequently, estimated sediment delivery rates must be evaluated in light of this fact. Given the research purposes of this study, however, the 1931 depiction provided a sufficiently realistic depiction of how land use, agricultural practices, and therefore sediment delivery rates to surface waters were different. The alternative would not entail much additional work. It would require amending the field and tract delineations to reflect changes in landuse and ownership and repeating the flow routing assignment on this coverage. Comparing the graphic output to current scenarios would be more visually demanding.

## Results

Tables 6 and 7 contain the results of spatial and statistical analyses of the WIN output for the five agricultural land use scenarios. Comparing the column depicting percent of total watershed acreage for the two scenarios illustrates how the acreage subject to non-conservation practices suffers a significant increase in sediment delivery. With conservation practices in place, the status quo in Dane County, 94.5 % of the watershed acreage generates less than 1 ton/acre/year of sediment. In contrast, scenario 2 results in more than half the watershed producing sediment at greater than 1 ton/acre/year. These results indicate that agricultural conservation can practices have dramatic impact on sediment delivery from cropped fields and are valid targets for improving environmental quality. These results also quantitatively document the benefits derived from the conservation provisions of the 1985 FSA and 1990 FACTA.

Table 7 compares data for all five scenarios aggregated at the subwatershed level. Both soil loss (USLE results) and sediment delivery results are expressed in tons/ year. Contributing cropped acreage has been factored into the summaries. Also, these results only reflect total sediment delivery from cropped fields, excluding non-cropped lands. Figure 3 illustrates the relative sediment contribution of all land cover types for scenario 1. One can see that cropped fields produce the majority of sediment. Figure 4 offers a spatial dimension to the WIN result for the first two scenarios.

Scenario 1, the current land use with agricultural conservation measures was used as a point of reference to compare scenarios 2 through 5. Implementation of conservation treatment due to federal programs results in a more than three-fold reduction in soil loss and sediment delivery. However, comparison of scenario 2 and 4 illustrates that much more could be achieved in pollutant reduction if all highly erodible cropped fields (A>T) were brought within tolerable limits.

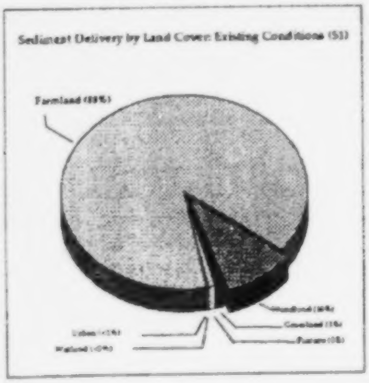


Figure 3

**Table 6: Sediment delivery rates in Garfoot Creek subwatershed for two land-use scenarios.**

| | Current Land-use with Conservation Measures | | Current Land-use – No Conservation Measures | |
|---|---|---|---|---|
| Sediment Delivery Rate in Tons/Acre/yr. | # of Fields | % of Total Watershed Acreage | # of Fields | % of Total Watershed Acreage |
| <1 | 222 | 94.5 | 109 | 45.0 |
| 1-3 | 6 | 2.5 | 26 | 9.5 |
| 3-10 | 1 | 0.7 | 56 | 26.5 |
| >10 | 2 | 2.3 | 40 | 19.0 |
| Totals: | 231 | 100 | 231 | 100 |

Changes in agricultural practices have less of an impact on estimated soil loss and sediment delivery than changes in landcover alone. Under scenario 3, reverting highly erodible cropped fields to forest cover (taking them out of production) would produce more water quality benefits than only additional conservation treatment (scenario 4). In all scenarios, sediment delivery was less than 20% of potential soil erosion estimated by WIN using the USLE. The first two scenarios exhibited higher sediment delivery as a percent of total soil loss as compared to the last three scenarios. This is probably due to two factors: (1) greater acreage of cropped fields in the study area; and (2) the type of fields contributing sediment. Fields closer to receiving waters tend to have higher sediment delivery rates because of increased surface water flow (contributing upland fields) and shorter runoff travel distances. This relationship, sug-gests that water quality impacts could be exaggerated if attention is focused only on USLE results. WIN attempts to factor in other variables influencing sediment transport including hydrology and topography thus providing a potentially more reliable and valid prediction of agricultural land use impacts on environmental quality. Of course, as with all models attempting to simulate the real world, accuracy and validity are a reflection of the quality of the model, the user, and the process of interpreting results.

## Conclusions

Water quality is just one of many measures of environmental quality. Similar spatial and statistical analyses can be performed using other measures. This would expand the information base used by public officials to assess the impacts of alternative agricultural land management policies. GIS/LIS techniques can enable rural resource planners and policy makers to more effectively and efficiently evaluate different natural and cultural variables. Existing tabular models can be supplemented using GIS/LIS by adding a spatial and temporal dimension to investigations. In a digital environment, resources can be interactively manipulated to create alternative scenarios to illustrate the relative impacts of policy and planning choices. Large geographic areas can be more easily evaluated to assess the impact of decisions affecting resource quantity, distribution and quality. From a water quality perspective, model input preparation using GIS/LIS saves hours of manual work that would otherwise be necessary. Such improvements in scale, breadth and rate of assessment are significant and important improvements in planning capabilities.

The statistical summary demonstrates the role that agricultural practices and forestry play in determining environmental quality. Converting all highly erodible cropped lands remaining after implementation of federal conservation requirements to forest cover, would provide the highest level of water quality protection in the study area. However, water quality is only one measure of environmental quality and these benefits must be realistically weighed against other land use objectives including crop production and/or wildlife management. Forests also serve important functions such as providing habitat for wildlife. When cumulative roles are compared, the relative contribution of forest cover and agricultural practices to environmental quality may balance differently.

## Acknowledgements

This work was funded by McIntire-Stennis Grant #WIS 3292 from the College of Agricultural and Life Sciences, School of Natural Resources, University of Wisconsin-Madison.

The authors would like to especially thank Eva Serra who developed much of the base data for this project as part of her habitat suitability research. Aman Tandias is acknowledged for development of the Fortran algorithms that helped automate the flow-routing process.

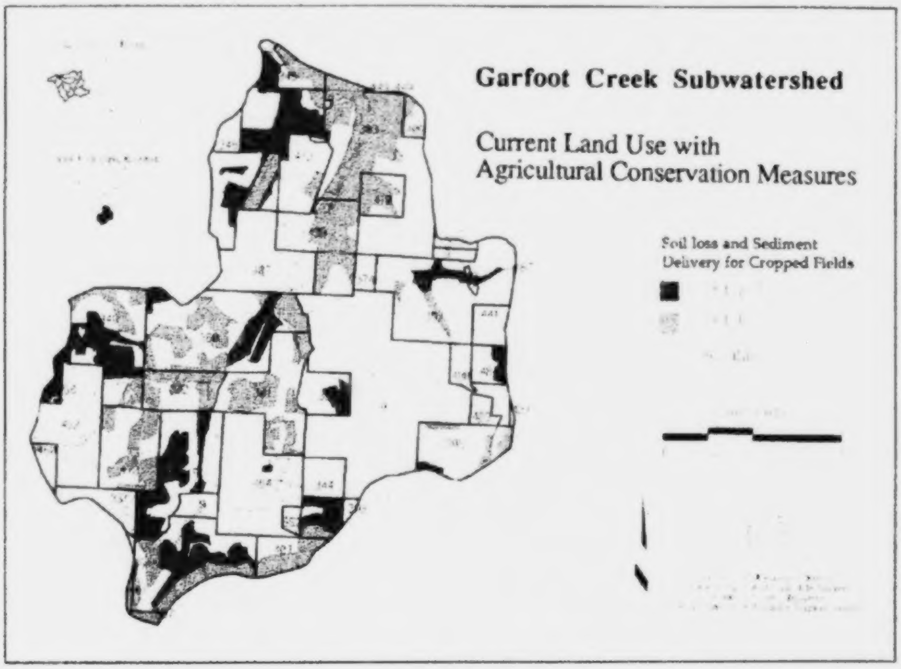

Garfoot Creek Subwatershed
Current Land Use with Agricultural Conservation Measures
Soil loss and Sediment Delivery for Cropped Fields

**Figure 4a**

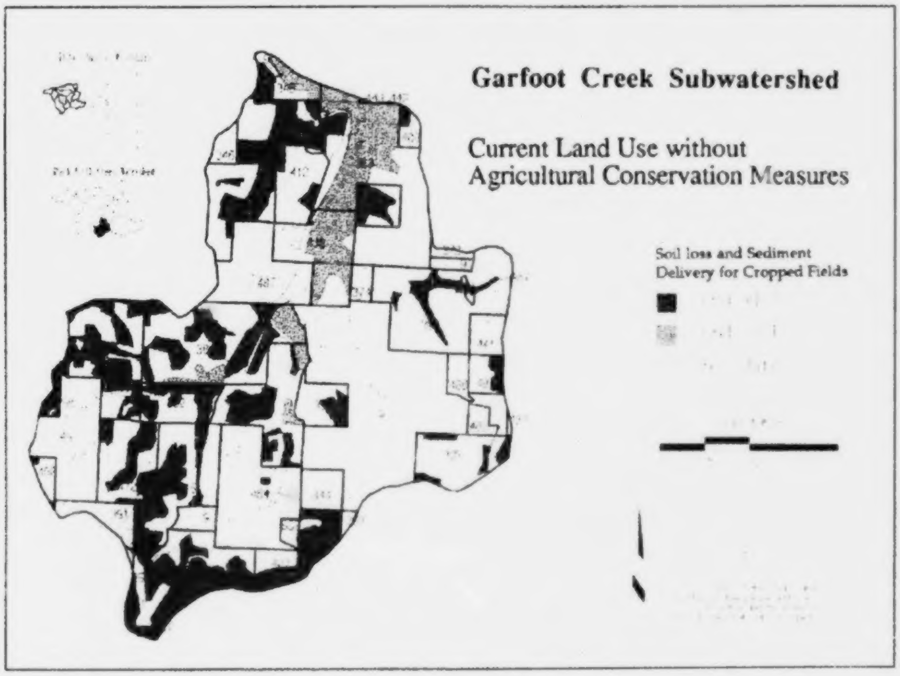

Garfoot Creek Subwatershed
Current Land Use without Agricultural Conservation Measures
Soil loss and Sediment Delivery for Cropped Fields

**Figure 4b**

**Table 7: Estimated soil loss and sediment delivery for agricultural landuse scenarios in Barfoot Creek subwatershed**

| Land-use Scenarios | Agricultural Acres | | Total USLE | | Sediment Delivered | | Sediment Delivered as % of Total Soil Loss |
|---|---|---|---|---|---|---|---|
| | Total | %Change | Tons/Yr. | %Change | Tons/Yr. | %Change | |
| **1991- With conservation measures** | 1,192 ac. | -- | 9,131 | -- | 1,680 | -- | 18% |
| **1991 - No conservation measures** | 1,192 ac. | 0% | 25,893 | +180% | 4,411 | +162% | 17% |
| **Where A>T revert to Forest** | 708 ac. | -40% | 2,183 | -76% | 238 | -86% | 11% |
| **Where A>T make A=T** | 1,192 ac. | 0% | 3,637 | -60% | 447 | -73% | 12% |
| **1931 - No conservation measures** | 1,562 ac. | +30% | 12,444 | +36% | 1,428 | -15% | 11% |

## References

Baun, Ken 1987, *The Wisconsin Nonpoint (WIN) Model User's Manual Version 2.2.* Publication No. WR-208-88. Wisconsin Department of Natural resources, Madison WI.

Baun, Ken 1988, *The Wisconsin Nonpoint (WIN) Model Application Demonstration Version 1.0* Publication No. WR-208-88. Wisconsin Department of Natural resources, Madison WI.

Bormann, F. H., G.E. Likens, T. G. Siccama, R.S. Pierce and J. S. Eaton. 1974. The Effect of Deforestation on Ecosystem Export and the Steady-state Condition of Hubbard Brook. *Ecol. Monograph*, 44(3): 255-257

Finley, R.W. 1951, The Original Vegetation Cover of Wisconsin, Ph.D. Dissertation, University of Wisconsin, Madison, Wisconsin.

Kaufer, N. 1976, Inventory of LEI Documents in Wisconsin Historical Archives. Unpublished Public Records Register. Wisconsin State Historical Society, Madison, Wisconsin.

Likens, G.E., F.H. Bormann, R.S. Pierce, J.S. Eaton, N.M. Johnson. 1977. *Biogeochemistry of a Forested Ecosystem.* Springer Verlag, NY, NY.

Schroeder, R. 1985, Habitat Suitability Index Models: Eastern Wild Turkey *Biological Report 82* (10.106). Fish and Wildlife Service, U.S. Department of Interior, Washington, DC.

Serra, Eva, Thum, P.G., Ventura, S.J., Niemann, B.J. Jr. 1990, Evaluating the Relationships Between Agricultural Forest Management and Environmental Quality Using a Multipurpose LIS, *1990 ASPRS/ACSM Annual Convention,* vol 3, pp.261-270.

# D1 Implementation and Impediments

**GIS two years after purchase (Where are we?)**

Craig Sharun, Canadian Forest Products Limited,
Grande Prairie, AB, Canada

**Distributing GIS data in a computer network**

Bert Veenendaal et al., Dept. of GIS, Curtin
University of Technology, Perth, Western Australia

**MIS and GIS - Can they work together?**

Bruce Abbott et al., Minnesota Department of
Natural Resources, St.Paul, MN, USA

**Trends in GIS technology. What are the impediments?**

Carl Reed, Genasys, Fort Collins, CO, USA

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# GIS two years after purchase (where are we?)

**Craig Sharun**

GIS Coordinator
Canadian Forest Products Limited
Alberta Operations
Postal Bag 100
Grande Prairie, AB
Canada T8V 3A3

## Abstract

*No matter where you go in the natural resources or municipal sectors these days, when the discussion of data storage or handling comes up, the topic eventually leads to Geographic Information Systems (GISs). There are a number of companies and organizations that have made the decision to go to a GIS or are planning on purchasing a GIS in the near future.*

*At CANFOR we have just recently passed this initial stage of GIS development. This presentation will cover the steps of GIS development that we found necessary to complete in order for us to get our GIS up and running. It will also cover some of the applications that the utilization of GIS technology has presented to our organization that would not have been possible without the installation of a GIS.*

## Introduction

So you have just shelled out $30,000.00 for a GIS software package and another $50,000.00 for the necessary hardware to run this package. You are now the proud owner of a GIS system.

Now the fun begins!

The first step in getting your GIS operational is to select someone to operate the system. Now, would you tell a medical student to study medicine for an hour or two once a week, when he can find the time, and expect this student to become a doctor? This approach to operating a GIS will more than likely produce the same results (or lack of them) as the medical student scenario. GIS software packages are very complex systems that represent leading edge technology and all of the spectacular achievements and dismal failures associated with it. The nature of this type of technology requires that you dedicate an individual on a full time basis to design and develop your GIS applications.

GIS systems are not only difficult to utilize but they are rapidly changing. For example, since my arrival at CANFOR fourteen months ago, I have installed FIVE different upgraded versions of our GIS software. The new versions have included changes that appear transparent, up to a complete overhaul of the software which included a radical change in the data structure and storage concepts. This is strong enough reason to dedicate someone on a full time basis just so that the operator can keep abreast of the changes in the technology.

One other thing to keep in mind when selecting the person to operate your new GIS is to try to select an individual that has a strong background in the area that the system will be applied. You will find that someone who fully understands the required output of your GIS will develop applications that will meet these requirements and not what his or her understanding of the application is.

Now that you have a GIS wizard in place to operate your GIS, you need data to work with. If you thought that the cost of hard and software for your GIS was pricey, you're not going to like this part. On average the cost of the GIS system is relatively small in relation to the cost of data you need to obtain valid results.

## Data acquisition and loading

Where will your data come from? How will you load your data? In Alberta, the Government is presently upgrading its Forest Inventory on to a 1:20000 scale base map. This map base is available to the public in a digital format at a very reasonable cost. It does not include any forest cover information for lands contained within Forest Management Areas (FMA's) which now cover approximately 80% of the forested area of the province. Since we manage an FMA we are responsible to provide our own forest cover data.

The method that we have selected consists of hand digitizing our existing forest inventory maps on to the new base map. The reasons for utilizing this method stem from the fact that our existing forest inventory maps do not overlay exactly on top of the 1:20000 base maps. When we attempted this we found, in some instances, that the inventory maps would have rivers located half way up hill sides or lakes would have mature pine stands growing inside them. We felt this situation could best be handled through manual digitizing so that the inventory maps could be shifted to match geographical features.

## Applications

Well now you have everything in place to start utilizing your system, almost! One of the first things you will find out about a GIS is it allows you to utilize and display more data than was possible with a non-digital information base. Soon you will have compiled more data than you previously thought possible. All this extra data requires new methods of analysis and new insight into how you should be managing your landbase. This requires you design applications to meet these needs and allow flexibility to adapt to future demands. This is where the addition of a person with a strong background in the area where you are applying your GIS will help tremendously. If the operator of your system understands what is required from the GIS and can foresee possible future demands, you can expect promising results for your GIS.

Once all your data has been loaded into the GIS and you are well on your way to building GIS applications, a number of options that were not available prior to having a digital data base become readily available. The first option that becomes apparent is data is easily modified and updated. This is a curse as well as a blessing. Being able to rapidly update your spatial and non spatial data is probably one of the quickest paybacks your GIS will avail to you but now that this process is so easy everyone wants everything changed.

At this time it should be mentioned the updates do progress rapidly as long as you are not required to produce hard copy updates. At CANFOR this is the situation we are in. Our maps are reproduced a number of times a week by members of the woodlands staff for field maps and other applications. For this reason we plot all of our maps on a polyester film. The maps are then reproduced on a blue-print machine. Our present map sheets require approximately three to four hours to plot using a pen plotter. This produces a bottle neck in our otherwise rapidupdate process. One solution to this problem is to purchase a colour electrostatic plotter. This solution will plot our maps in black and white in approximately one minute or in colour in under eight minutes. The drawback to this solution is colour electro-static plotters cost between $50,000.00 and $60,000.00. With the economy being what it is, trying to convince management to authorize a $60,000.00 purchase is somewhat arduous. Another solution is to eliminate the need to produce hard copy maps and establish a method to exchange digital map products with the required government agencies.

One way to reduce the initial cost of data collection for your GIS, or at least diversify the expense, is to gather data in selected key areas and then develop site specific applications for these areas. This is the approach we have adopted at CANFOR. We started by acquiring data for approximately 1/3 of our operating area (200,000 hectares). We then began to utilize this data in applications contained within this area. For example we have developed a landuse billing application we utilize for generating a report to calculates timber damages which result from seismic or pipeline work within the FMA. To use this application, we load the GIS with the location of the seismic line or pipeline then generate a buffer around the linear object corresponding to the width of the disturbance. This buffer is then overlaid on to our forest cover spatial data base. This process generates a table which lists the type of timber and area of the stand affected by the linear disturbance. This information is then loaded into a DBASE program that references a lookup table to calculate the monetary value of the damaged timber.

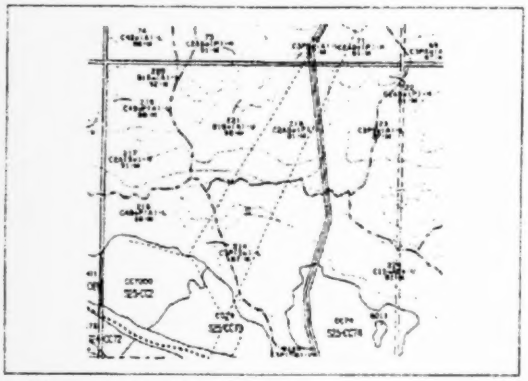


Figure 1 Landuse Application

A variation on this application was to generate a grid which corresponds to a township grid for Alberta. This grid was broken down to a legal subdivision level. This application was developed to provide a detailed legal description for all of our Licence of Occupation applications on our main roads. Once again, the proposed road location is digitized on to the grid and, when an overlay function is selected, the GIS generates a report listing off the legal subdivisions the road passes through. The linework loaded into the system is broken at the township boundaries so when the overlay process is initiated, the linework can be separated by township to allow for identification of the different townships and ranges.

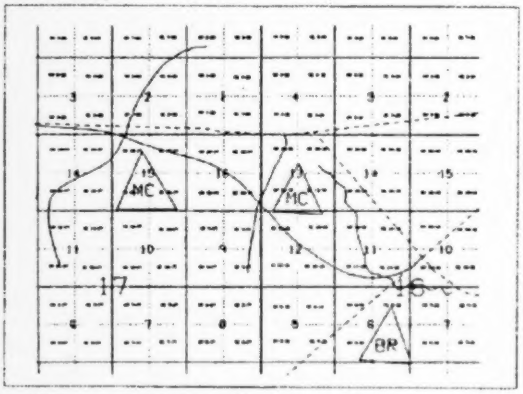


Figure 2 L.O.C. Application

We have also used a site specific approach to develop Annual Operating Plan maps (AOP) maps. These maps are a detailed plan of where and how we are planning to harvest timber over the next operating year and a preliminary plan for where operations will be conducted over the next four years. We use the drafting functions of the GIS to generate the harvest maps. The maps are originally developed on aerial photos. The ability to register these photos to our digital map base with the GIS is a major asset. It not only saves time but it increases the accuracy of the final product. Once the information has been transferred from the photos, the analytical functions of the GIS are used to generate the areas for the cutblocks. Before acquiring our GIS these maps, once completed via hand drafting on Dylars, were hand coloured to distinguish between the first and second pass harvest areas. This task was labour intensive due to the fact seven copies of our AOP were required, each containing approximately 15 maps. The use of the GIS has enabled us to create different shading patterns of the first and second pass harvest areas. These patterns can be reproduced on our blueprint machine and were accepted by the Alberta Forest Service.

One of the more extensive areas we are gathering data is in silviculture. When we started keeping silviculture records they were stored in the traditional method of "lots of filing cabinets". This system worked but required vast amounts of time for someone to access the data. The information was periodically accessed by the silviculture forester but its main use was to calculate our annual reforestation liability. It was obvious that the best way to handle this information was to store it in a computerized record system. Our answer was to create a DBASE program that would store and organize the data as well as generating the annual reforestation liability report data (See Figure 3). This was all put in place prior to the acquisition of our GIS. The silvicultural record system, once computerized, was a vast improvement over the old method but it still required the human link between the data base and the maps. The next logical step was to tie the silviculture records to the map base with the GIS.

This turned out to be a very simple task due to the fact that the records were stored in DBASE which is the database used to drive our GIS. The records were stored with a unique identifier by including, in the cut block number, a prefix that indicated the location of the block. There were a few exceptions to this rule in some cutblocks that were split by a road or pipeline and had to be identified in the GIS as separate entities. The silviculture records for this block indicated both sections of the block were treated as one.

Once the link between the spatial and nonspatial data has been established it allows us to generate maps displaying areas which have passed regeneration surveys or are still awaiting to be surveyed, maps showing areas requiring reforestation treatments over the next five years and the type of treatment required in each area. Maps can be generated to display any attribute or combination of attributes contained in the silvicultural database.

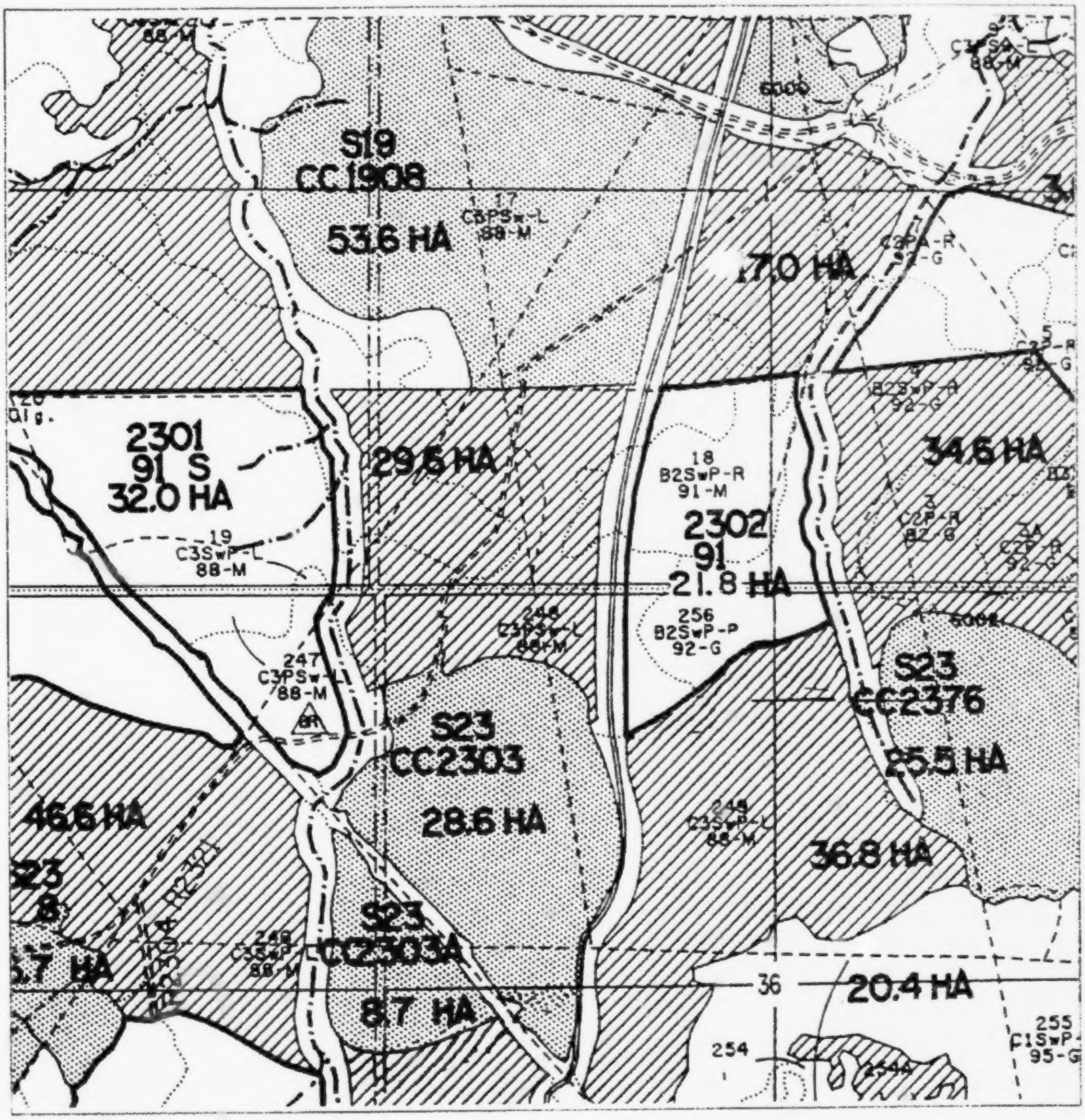

S19
CC1908
53.6 HA
C3PSw-L
88-M
17.0 HA
C3PA-R
92-G
2301
91 S
32.0 HA
29.6 HA
19
C3SwP-L
88-M
18
B2SwP-R
91-M
2302
91
21.8 HA
34.6 HA
B2SwP-R
92-G
C2P-R
92-G
256
B2SwP-P
92-G
248
C3PSw-L
88-M
247
C3PSw-L
88-M
S23
CC2303
28.6 HA
S23
CC2376
25.5 HA
46.6 HA
249
C3SwP-L
88-M
36.8 HA
S23
CC2303A
248
C3SwP
88-M
8.7 HA
36
20.4 HA
254
255
C1SwP
95-G
6000

Figure 3 Sample of a Digital AOP Map

## New technologies

The establishment of a GIS in the work place will provide a stepping stone to a number of advanced technologies. The basis for these advances will be the digital map base that drives the GIS. Once the spatial information from your maps is in a digital form, data can be imported or exported easily.

One of the first methods of importing digital information into our database we pursued was satellite imagery. The area of Alberta that we manage (our FMA) consists of approximately 655,000 hectares. Because of the size of this area it is difficult and expensive to gather and update information. The information we require can be classified into two areas; the first being accurate forest cover information, the second an accurate and current update of linear disturbances and harvesting activity within the FMA. We did not feel comfortable present satellite imagery technology could provide us with the degree of accuracy we require for forest cover information. This information must supply tree heights with plus or minus 1 meter accuracy, species identification, as well as understory species and height. We did feel though that satellite imagery would work well to define linear disturbances or harvest areas. We looked into this application and found satellite imagery could be provided to us in a digital format which could be imported into our GIS in the form of vectors. This appeared to be the best way to obtain this information until the price was discussed. We found it hard to believe we were only purchasing the images from the satellite not the entire satellite. Keep in mind though this was a complete package ready to be imported into our GIS and this was just one contractor. We did not proceed with this application much past this point as we felt this option was cost prohibitive. We are looking into more cost effective applications for this technology.

A second method of gathering data for our spatial data base also utilized satellite information. This information was provided through a Global Positioning System (GPS).

It is very important, if not essential when working in the forest to maintain an accurate idea of your location when developing new areas of virgin forest. Most areas in Alberta have an established network of seismic lines which allow access and provide tie points when working in the bush. The detailed work associated with proper forest management requires a large degree of reconnaissance work. Areas that fall into this category would be cutblock layout and road layout. The base method of cut block layout at our operation in Grande Prairie is irregular patch layout or timber type layout. This method does not lend itself to compass and chaining through the bush and consequently creates positional inaccuracies. Through the use of a GPS these positional inaccuracies are greatly reduced or eliminated. In short this technology allows the field staff to concentrate on the job at hand instead of maintaining a positional fix of their location.

A brief description of how a GPS works may be warranted. This technology utilized an array of satellites controlled by the United States Department of Defence. The present array consists of 18 satellites with a full complement of 21 operational and three spares to be in place by 1993. To keep it simple a receiver on the ground picks up signals sent by the satellites. These signals contain positional information (longitude and latitude), elevation and time. The receiver picks up the signals and by means of triangulating between the satellite and the receiver a ground position is established. A signal must be received from at least three different satellites at one time in order for a position to be calculated. This type of fix is called a 2D-FIX. If a signal is received from four or more satellites at one time then a 3D-FIX is established. The difference between these two types of fixes is a 2D-FIX which only gives a horizontal position value and is less accurate than the 3D-FIX which provides a horizontal and vertical positional fix.

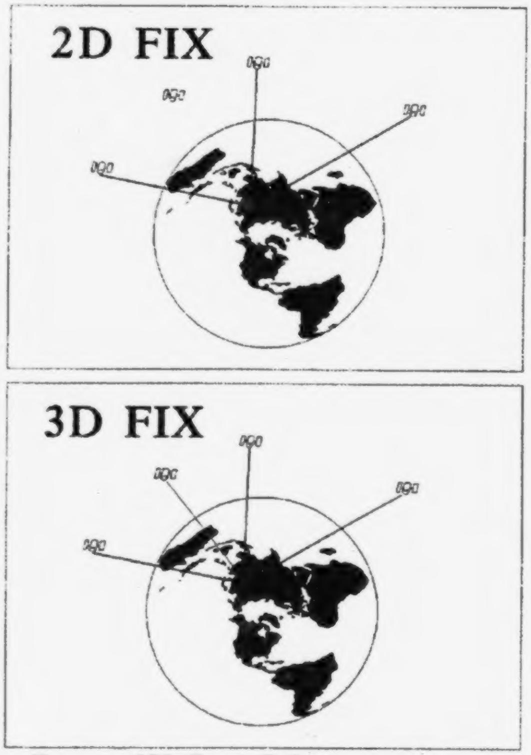


Figure 4 GPS 2D and 3D Fixes

Now to get to the problem at hand. At our operations in Grande Prairie we base our future reforestation costs and harvest volumes on the areas of the proposed cutblocks for this operating year. The only way to confirm this was to wait until all of the area had been harvested then fly it and update the block areas and shapes from the new photography. It was noted there were a lot of discrepancies between the proposed areas and shapes of the blocks and the final results. We felt the best way to reduce or eliminate these discrepancies was to get a better handle on where the block boundaries were in relation to where the maps showed them to be. The solution seemed to be the use of a GPS. But would it work in our present application?

To answer this question a trial was undertaken to see if a GPS could operate under the forest canopy. This would allow us to use the GPS during block layout thus substantially reducing the additional cost associated with using the GPS. The trial consisted of gathering data with a single channel and a six channel GPS receiver under three different forest cover types. It became apparent quite quickly during the field trials that under forest canopy the single channel receiver was not going to provide adequate information. On the other hand the six channel receiver was able to provide a strong 2D-FIX and a 3D-FIX as long as the satellites that were available maintained an azimuth of approximately 45 degrees or greater relative to the horizon. The six channel receiver also allowed us to gather data in a stream mode while walking through the bush, a mode not possible with the single channel receiver.

Once the data had been collected it was down loaded into the GPS software then converted to a form that could be imported into our GIS. The results were very favourable. The data showed some degree of variability, but over all a distinct path around the cutblock could be established. The data collected in all three timber types indicated that the signals could be collected under various timber types without too much difficulty. There were times when the signal was blocked but these incidents were few in all cases as long as three or more satellites remained visible and maintained an angle above the horizon of approximately 45 degrees.

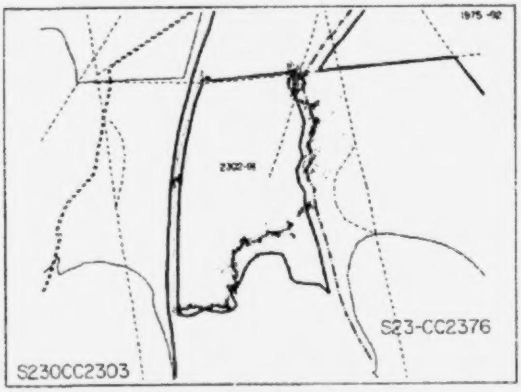


Figure 5 GPS Raw Data Overlay

So far I have covered some of the applications that a GIS can offer but you may have noticed most of the applications required very little analytical processing. This situation I feel is only due to the method of development we have chosen and does not indicate it will take other operations this amount of time to begin to fully utilize the strength of the analytical processing of a GIS. We have directed our efforts into areas that answer our most current and urgent needs. This, coupled with an incomplete data base, have led us to develop applications requiring minimal analytical processing.

Applications that do require large amounts of processing are currently being developed. One such project is Ecological Site Classification, better known in B.C. as Biogeoclimatic Zoning. This application will require the use of multithematic overlays which will analyze such information as slope, aspect, forest cover, soil, shrub, herb and elevation information.

Up to this point our GIS operation has been a one man show. By this, I mean one person has been intrusted to load the necessary data and develop various GIS applications identified by the members of our management team. Now all of the GIS applications are ready to be placed into the hands of the people who can best utilize the information, the management team. This can be addressed through the establishment of a Local Area Network (LAN). If you do not already have a LAN in place this becomes another unexpected cost in the establishment of a GIS. You will not realize the benefits of your GIS if it is not available to the people who can utilize it. We had planned to install a six station LAN in our offices in January, but due to financial restraints we have had to compromise and will be using a "moccasin" network between two GIS installations until it is possible to proceed with our original plan.

## Conclusion

GIS technology is no longer viewed as a surrealistic solution to handling spatial data. It is a technology being applied in various operations around the world in applications that range from a dentist mapping his patients' teeth to a convenience store owner looking to see the benefits of expansion. The application of GIS's are only limited by the people using them. If you have not already bought into the GIS solution, I feel you will eventually have to. The competitive edge provided through the use of GIS's will force you to keep up.

Once you have chosen to get into a GIS, keep a realistic set of objectives. The time frame I have laid out reflects the steps taken at CANFOR's operation in Grande Prairie. In this case the operator hired to use the system had three years' prior experience utilizing the software package that was in place in our operation. This eliminated the need for any training andsubsequently any delay in the start of our GIS. The learning curve for a GIS is very steep and experienced personnel are hard to find. Unless you have a person already in place, this will no doubt delay the initial stages of your GIS implementation.

Another factor which will dictate how fast you will progress is the availability and accuracy of the data you require. If data is presently available, you have an excellent start. If not, the various methods of data collection available will pose a problem as involved as the methods of initializing your GIS.

Take your time and develop applications that provide the best utilization of the power and flexibility of your GIS. Keep these applications flexible enough to be able to adapt to future changes in technology. Design your data base so it provides optimum availability to your data. It is always easier to group data within your GIS than it is to separate it.

Last but not least, let your GIS progress and grow at its own pace. A GIS has been compared to a child. If you force it along and demand it do something before it is ready, the results may not be favourable, but if you nurture it along and provide support and guidance it will develop into something that you can be proud of.

GIS'92

Symposium

Vancouver
British Columbia
February 1992

# Distributing GIS data in a computer network

**Bert Veenendaal and Doug Hudson**

Department of Geographic Information Systems
School of Computing Science
Curtin University of Technology
GPO Box U1987
Perth, Western Australia 6001

## Abstract

*Increasingly large volumes of geographic data are being accumulated and the use of computer communications networks is escalating. This results in an increased need to distribute and manage the available resources. Configuration of the computer network, the partitioning, replication and update of datasets, and the placement of both data and programs within the network are issues which must be addressed when designing a distributed Geographic Information System (GIS). The objective is to effectively utilize processing and storage resources and provide users with reliable systems that satisfy access and performance requirements. Initial design as well as ongoing management are factors to be considered. The problem is finding a solution without making too many restrictive assumptions. This paper discusses these distribution issues and their implications for GIS and provides some directions for further research.*

## Introduction

"Working Smarter", the theme of this symposium, includes using the currently existing and developing technologies to the advantage of the users. Geographic Information Systems (GIS) are about integrating various technologies that can operate on geographic data, with the objective of allowing the users to efficiently and effectively carry out their application specific tasks in the natural resources, environmental, land use management, planning, or other related areas. Distributed systems are yet another technology which can be integrated in a GIS environment to provide users with access to more data, quicker processing, and access to other resources.

Especially because of the nature of GIS data and the increasing usage of computer communications networks, distributed systems will play a prominent role in the growth and effectiveness of GIS systems. The large volumes of spatially-related data being accumulated for large geographic regions, the increasing reliance on intra- and inter-organisational interactions and exchanges, and the resulting complex data management and control requirements are preparing the way for a new generation of GIS, namely Distributed GIS.

Integration of the GIS and communication network technologies raises the following issues concerning the distribution of geographic data and resources on a computer network. What network configuration is required to satisfy the needs of a distributed GIS? What is the best way to partition the geographic database so that it can be placed on the network? Where in the network is the data to be placed so that the given constraints are adhered to?

To what extent can the data be replicated across the network, and how are updates facilitated? Where are the programs located and where will the processing take place?

## GIS data distribution design

Data distribution design is part of an overall database design and is concerned with distributing the data throughout a computer communications network according to given requirements and constraints so that certain objectives can be achieved (Figure 1).

This distribution entails the partitioning of a database into units that can be distributed onto the network sites, replication of these partitions in the network, and placement of these data units at particular sites. These three, partitioning, replication and placement, are closely interrelated and influence each other (Ozsu and Valduriez, 1991). Any decisions made with one of these aspects of data distribution affects the others.

Data migration is concerned with the ongoing relocation of the data to ensure the objectives are continually met since the requirements and constraints may vary dynamically over time.

An objective in a distributed GIS is to ensure that the data is accessible to the user when and where it is required. Often data is available but not immediately accessible because it is physically located elsewhere or it may be in a form that is inaccessible. Making data accessible may involve placing the original data on a closer site or replicating the data at or near network sites where it is heavily accessed.

Closely related to accessibility is data reliability. If the data is replicated at multiple sites in a computer network and a site fails or becomes inaccessible through a communications failure, even though some of the data may not be accessed, the user may be able to carry on with the replicated versions.

Another objective of data distribution design is to improve performance. Locating the data proximate to users may decrease the time required for database accesses, decrease the contention for storage and processing resources by distributing the load, and improve query response times (Ozsu and Valduriez, 1991a).

## Factors affecting data distribution

Data distribution design is influenced by a large number of factors. These include organizational, data, and user requirements as well as system constraints which are based on storage, processing and communications capacities at each site in the network. In addition, the network configuration itself heavily influences where the data is to be placed on a network so as to satisfy the objectives.

Network configuration involves determining the location of nodes, the facilities provided at each node, the interconnection of the nodes by links, and the capacities of the links. Designing an adequate configuration must take into account the geographic locations of the users, the performance and capacity requirements of each user at each site, and of course the costs involved in implementing and maintaining the network. In many cases the communications network may already be defined and in place and therefore the configuration may already be predetermined. Making efficient use of existing links and planning for link upgrades will require knowing the traffic loads that GIS users will incur.

The configuration required depends largely on the organizational structure and the geographic locations of the GIS facilities. In the past, management and security requirements have dictated that the facilities remain central and centrally-controlled. Distributed technology allows facilities to be geographically dispersed whereas the control can either remain centralized or can become decentralized if so desired.

An organization with separate departments housed within a single building or small group of buildings would typically use a Local Area Network (LAN) to interconnect the various computing resources. Where regional offices are geographically dispersed a Metropolitan Area Network (MAN) or Wide Area Network (WAN) may be required. In most cases, centralized control of the data and other facilities is desirable. Decentralized control would be required if separate organisations were to be interconnected; even though these organizations may be able to exchange data and other resources, they would individually want to retain control of their own datasets or at least determine who has control as in the case of a data custodian.

In addition to knowing the network interconnection, the resources required at each node must be determined. User and GIS requirements will indicate the processing hardware and software, memory and disk storage, and peripheral resources such as digitizers, scanners, and plotters as well as the communications facilities required at each node. Most GIS have considerable storage and processing requirements which place an extra burden on communication facilities in a distributed environment. The processing, storage, and communications channel capacities will, in turn, affect how and what data is to be placed at each site.

Organizational and security requirements determine who has access to what data and how that data is to be used. This may limit the extent to which data may be distributed in a distributed GIS. Which data is accessed, where is it accessed, and by whom is it required and for what purposes? Data usage might range from simple queries and retrievals by client users, through data capture and maintenance by data custodians, to complex data analysis and modelling by applications experts. Similar types of users accessing the same data may be scattered across the different sites in the network while many different types of users may be working from the same site and accessing many different databases located at various other sites in the network. Data may therefore have to be restricted to certain sites and with varying privileges to allow certain modes of access for the various types of users.

Data usage patterns will affect how the data is to be placed in a distributed GIS network. Data usage, as well as the location of the users, may change over time and may vary across the user groups. Such a dynamic situation involving data migration makes the problem of data distribution design extremely difficult.

In addition to the data, the applications programs must be optimally placed in a distributed GIS network. Obviously interrelated with how the data will be distributed, the placement of applications programs will depend on such factors as the location of the users and data, the data usage, the processing capacities (including memory and disk swap space), and the organisational and contractual arrangements in place (eg. software license agreements). The placement of programs also involves the types of queries, the processing order of the queries, and how the queries can be executed in parallel at different sites in the network (Wah, 1984).

Many of the problems in distributed database design have been discussed at length in the general literature and can be applied to the distribution of geographic data in a distributed GIS. It is, however, the spatial nature of geographic data that adds another dimension of complexity to these problems.

## Data partitioning

Before data can be placed at each site, the manner in which the data is to be partitioned must be considered. Rather than have all the data stored at one site, it can be partitioned into fragments which can then be distributed across the computer network.

Partitioning geographic data involves both the spatial and attribute (or non-spatial) components which must be kept closely linked in a GIS. Attribute tables can be horizontally partitioned into groups of records which satisfy a particular condition, or vertically partitioned by grouping attributes that have a high probability of being accessed together. Spatial data can be spatially partitioned by particular geographic regions (see Figure 2). In addition to the separate partitioning of the spatial or attribute components, the attribute data can also be partitioned from the spatial data. This results in effectively eight different combinations of partitioning the geographic data vertically, horizontally, and/or spatially.

Further partitioning can be obtained by separating the various attribute tables from each other or by separating the various spatial data layers from each other. Because the attribute tables are usually closely linked to the spatial data layers, this partitioning can be done so as to keep the spatial data layer and the corresponding attribute table together (Figure 3). Geographic data on a single site is traditionally already partitioned thematically into layers and often spatially into map sheets or tiles (Goodchild, 1990).

Most research on the data partitioning problem has been fairly recent and is, for the most part, limited to textual databases (Hevner and Rao, 1988). Solutions exist for doing either vertical partitioning or horizontal partitioning on textual databases. There is however no algorithm to fragment data both horizontally and vertically (Navathe and Ceri, 1985; Ozsu and Valduriez, 1991a). Most real-life partitioning would be mixed, and especially so with geographic databases where spatial partitioning can additionally be done. For example, it would be quite common to spatially partition a spatial data layer and to horizontally partition the related attribute table so that the spatial data and corresponding attribute data can remain together in the same partition (Figure 4).

Heuristic algorithms exist to vertically partition a dataset (Hevner and Rao, 1988; Ozsu and Valduriez, 1991b). Either a grouping method is used, which starts by assigning each attribute to a partition and then joins some of the partitions until some criteria are satisfied, or a splitting method is used, which starts with an attribute table and then decides how to partition the table based on the access behaviour of each attribute.

Hevner and Rao (1988) describe a splitting algorithm which first calculates an attribute affinity value for each pair of attributes indicating the importance of these two attributes appearing together in a partition. Secondly, a clustering algorithm is applied to rearrange the attributes by their affinity value. Finally these clusters are partitioned into subsets which have minimum interaction. This final partitioning step may or may not allow attributes to overlap the final partitions.

With horizontal partitioning, conditions are placed on one or more attributes to identify the records to be placed in each partition. This requires an understanding of the data and its usage in the distributed environment. Spatial partitioning can be done in a similar manner; the spatial components are derived by placing conditions on one or more corresponding attributes in the attribute table. Geographic data can be both spatially and horizontally partitioned to keep the spatial and attribute data together in the same partition.

## Data replication

The problem of distributed data replication is complex, especially when both spatial and attribute data are considered. The objectives of replication are to ensure reliability, increase availability and improve system performance by decreasing response time and providing parallelism by allowing simultaneous access to different copies of the same data (Hevner and Rao, 1988). A disadvantage of replication is the updating that is required for each replicated copy as well as the related problems of providing concurrency control, integrity of data, recovery from system failures and transaction abortions, and measures of security (Hevner and Rao, 1988).

A data replication strategy will generally fall somewhere between the one extreme of no data replication at all and the other extreme of full replication where a copy of all data is found at every site. Typically data that is critical for system reliability or data that is highly used from various network sites will be replicated while data with strict security requirements or data that is extremely volatile and requires frequent updating will not be replicated.

Several strategies including simulation and mathematical models can be used for evaluating data replication. Hevner and Rao (1988) describe a simple decision model which first places a data partition on a site based on the frequency of use of an application program at that site and the probability of that partition being accessed by the application. A greedy heuristic then finds additional sites on which to place copies of the data by determining the difference in the number of retrieval accesses that become local and the number of additional remote update accesses required to maintain that copy.

Distributed data replication is closely related to data partitioning for geographic data. It is, of course, the partitions that may have to be replicated. In addition, the linkage between the spatial and attribute components of the data must be maintained. This may mean, for example, that if an attribute table is vertically partitioned, the spatial data may need to be replicated to maintain the linkage of the two attribute partitions to the same spatial data (Figure 5). Likewise, if the spatial data is partitioned, the attribute data may have to be replicated. For example, in Figure 6 the soils layer is spatially partitioned and the attribute table is replicated so that it resides at the same site and is linked to the spatial data.

An additional aspect of replication that has particular importance in GIS is the potential need to store multiple representations of spatial data in order to accommodate requests for data at various scales. In a highly active user environment it may prove computationally too expensive to repeatedly derive small scale maps or other highly abstracted information products from a detailed set of base data. In these situations the database/network administrators may decide to precompute some results based on expected demand patterns. This may take the form of one or more complete derivations of a data coverage at commonly requested scales including coordinate transformation, feature generalization, and symbol modification.

Although not exact replicas of the base data, multiple representations certainly qualify as redundant data and require an update strategy of their own to maintain database consistency. Maintaining this consistency is now much more than simply transferring modified values to new locations; it may require a substantial recomputation of values in addition to the data transfers.

To further complicate matters, it may be necessary to develop a partial multiple representation strategy. High usage of a particular abstraction of data (for example, a request for a map at a scale smaller than that of the base data) would indicate that the abstraction should be precomputed and stored at one or more sites. However, frequent base data updates may negate these benefits because of high update propagation costs. By careful analysis of the procedures used in the abstraction process, it may be possible to partially derive these abstractions in such a way as to minimize the costs of the remaining computation and communication in both query and update processes (Hudson, 1989).

These multiple representations, either full or partial, may be subjected to the same partitioning and site replication strategies as their single representation counterparts bringing a further measure of complexity into the analyses. Mathematical models to determine replication evaluation strategies are part of an active research program at Curtin University.

## Data placement

Distributed data placement is concerned with what data will be allocated to the various sites in the computer network. This is the area where most of the previous research in data distribution design has been conducted. While early data placement models assumed the unit of data distribution to be the file, later models generally handle any partition. A survey of these models can be found in Hevner and Rao (1988). The file allocation problem is discussed by Wah (1984) while the partition or database allocation problem differences and solution strategies are discussed by Ozsu and Valduriez (1991b). In distributed GIS, the unit of data distribution can be a geographic layer or more specifically the spatial and/or attribute data partitions.

Many of the data allocation models measure the effectiveness of the data placement strategies using storage, retrieval and update costs. These costs can be determined from such factors as storage capacity, processing capacity, type of data usage (eg. retrieve data or update data), frequency of data access, and communications capacities. The placement strategies attempt to meet performance objectives by minimizing response times and maximizing system throughputs at each site. Especially with the large volumes of data and processing intensive operations typical of GIS, it is important to know the effects of placing data at particular sites in a network if efficiency and effectiveness of GIS are to be maintained or enhanced.

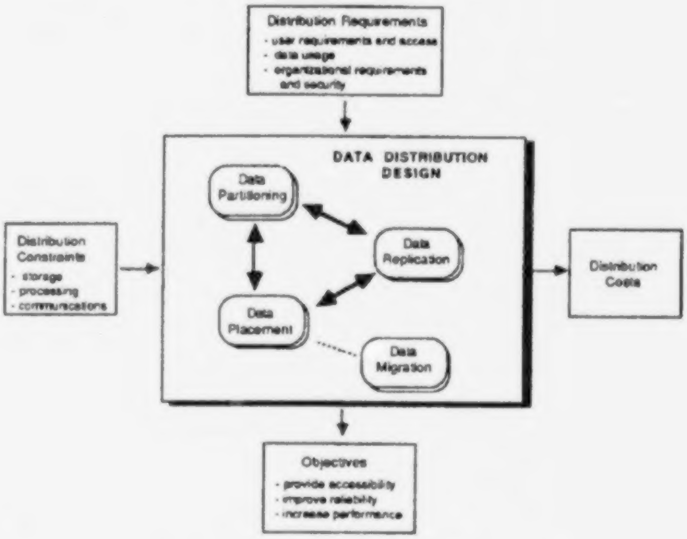

Distribution Requirements
- user requirements and access
- data usage
- organizational requirements and security
DATA DISTRIBUTION DESIGN
Data Partitioning
Data Replication
Data Placement
Data Migration
Distribution Constraints
- storage
- processing
- communications
Distribution Costs
Objectives
- provide accessibility
- improve reliability
- increase performance

FIGURE 1 - Data Distribution Design Problem

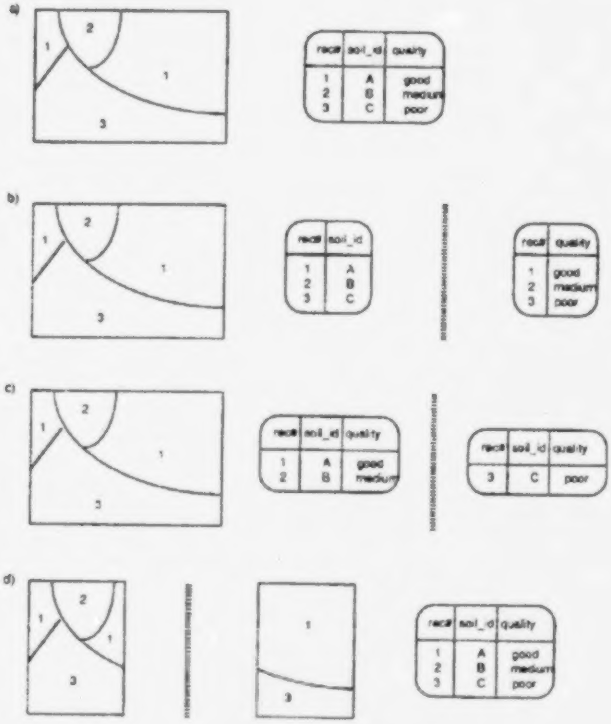

a)
1
2
1
3
rec# soil_id quality
1 A good
2 B medium
3 C poor
b)
rec# soil_id
1 A
2 B
3 C
rec# quality
1 good
2 medium
3 poor
c)
rec# soil_id quality
1 A good
2 B medium
rec# soil_id quality
3 C poor
d)
1
2
1
3
1
3
rec# soil_id quality
1 A good
2 B medium
3 C poor

FIGURE 2 - Types of Data Partitioning. a) the original geographic data b) vertical partitioning c) horizontal partitioning d) spatial partitioning

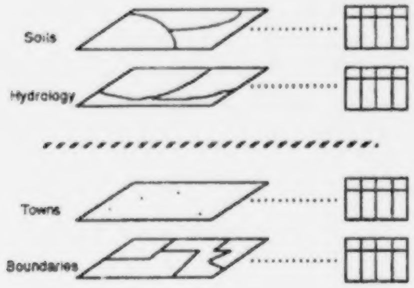

Soils
Hydrology
Towns
Boundaries

FIGURE 3 - Partitioning the spatial data layers and corresponding attribute tables

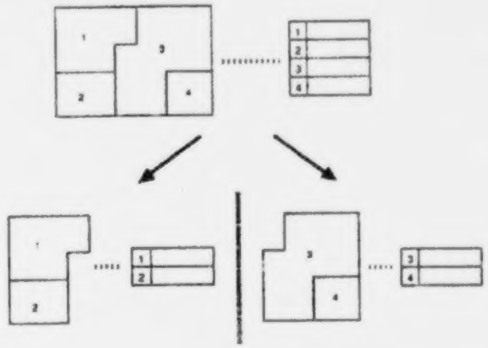

1
3
2
4
1
2
3
4
1
2
1
2
3
4
3
4

FIGURE 4 - Spatial partitioning of the spatial data and horizontal partitioning of the attribute data to retain spatial-attribute linkage

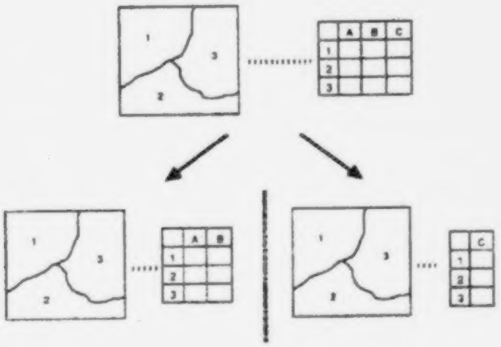

1
3
2
A
B
C
1
2
3
A
B
C

FIGURE 5 - Vertical partitioning of the attribute data and replication of the associated spatial data

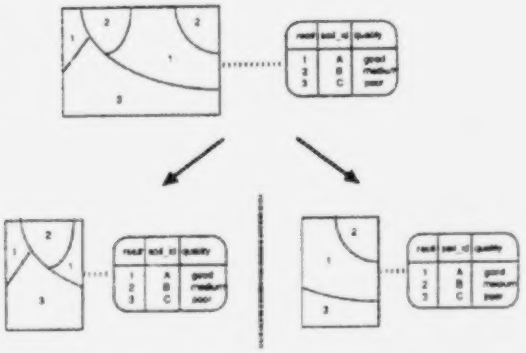

2
2
1
1
3
good
medium
poor

FIGURE 6 - Partitioning of the spatial data and replication of the attribute data

Data migration is essentially the dynamic version of data placement. Data usage at each site as well as the types of users at each site may change over time. There may be a locality of access for a dataset which may be time-varying. Geographic datasets themselves may evolve and grow as more data and new types of data are accumulated. The requirements and constraints of a distributed GIS may change as well as the costs due to upgrading with improved or new technology. In such a dynamic environment, the effects of an initial data distribution design may soon become lost. It may then be more efficient to allow the data to migrate to different sites from where it can be accessed.

Data migration will involve redoing the data placement step at occasional or regular intervals. This will undoubtably involve addressing the data partitioning and replication design steps again to reduce or improve the locality of access. The migration problem is further discussed by Wah (1984). Data distribution design is in effect not a once-off initial step in the overall system design phase, but an ongoing requirement for ensuring that the various distributed GIS design requirements are adhered to and the objectives are obtained.

## Conclusion

Distributed systems have received much attention by researchers and will continue to do so with the use of improved computer communications networks, faster hardware, and more efficient software techniques. Application requirements and user demands will also drive the need to further develop distributed systems. Even though much work has already been done in the areas of distributed processing, distributed operating systems and distributed databases, as well as in developing distributed systems, much remains to be done to make these systems fully operational for applications using GIS technology (Veenendaal, 1990).

Not only must the technologies become further advanced, but also the methodologies and procedures in using these technologies to produce efficient and effective distributed systems. Just as database design is important for GIS, so distributed database design, including data distribution, is important for distributed GIS. Many solutions and strategies have been developed for partitioning, replicating, and placing non-spatial data, but much more needs to be done in developing and integrating these techniques for use with spatial data.

Understanding the complexities of geographic data and their inter-relationships as well as the spatial operations performed on the data will be important in determining effective geographic data distribution design strategies, thereby providing users with the means of better understanding, managing, and most importantly using their most valuable resource - the data.

## References

Goodchild, M.F. (1990) Tiling Large Geographical Databases, Proceedings of First Symposium on the Design and Implementation of Large Spatial Databases (SSD '89), Springer-Verlag.

Hevner, A.R. and A. Rao (1988) Distributed Data Allocation Strategies, Advances in Computers, Volume 27.

Hudson, D.L. (1989) Autonomous View-Stages: Materialized Support for View Update Propagation, Ph.D. Thesis, Department of Surveying Engineering, University of Maine, Orono, ME.

Navathe, S.B. and S. Ceri (1985) A Comprehensive Approach to Fragmentation and Allocation of Data in Distributed Databases, Tutorial: distributed Database Management (eds. J.A. Larson and S. Rahimi), IEEE Computer Society.

Ozsu, M.T. and P. Valduriez (1991a) Distributed Database Systems: Where Are We Now?, Computer, 24(8): 68-78, IEEE, August.

Ozsu, M.T. and P. Valduriez (1991b) Principles of Distributed Database Systems, Prentice Hall, Englewood Cliffs, New Jersey.

Veenendaal, B. (1990) Sharing GIS Data in a Networked Environment, Proceedings of AURISA 90, Canberra, Australia, November.

Wah, B. (1984) File Placement on Distributed Computer Systems, Computer, 17(1): 23-32, IEEE, January.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# MIS and GIS — Can they work together?

**Bruce W. Abbott**

Minnesota Department of Natural Resources
Division of Forestry
500 Lafayette Road
St. Paul, MN 55155-4044

**Christopher J. Cialek**

Minnesota Department of Administration
Land Management Information Center
658 Cedar Street
St. Paul, MN 55155-1603

## Abstract

*The roots and evolution of management information systems (MIS) and geographic information systems (GIS) are quite diverse and seemingly unrelated. MIS processes trace their roots to traditional accounting practices. They frequently begin with a detailed study of system requirements from which are developed 'bullet-proof' tabular database maintenance routines and efficient reporting of repetitive business functions. GIS, in contrast, has more recently developed, spurred on by rapid advances in computer-assisted graphics and spatial data processing. The GIS process typically begins with project-oriented, non-routine data entry, accesses data through a prototyping approach, and moves quickly to ad hoc data analysis and spatial modeling.*

*Clearly, MIS and GIS have evolved from two very different directions. However, the two disciplines appear to have converging interests in their emerging views of data as an invaluable resource. There is an increasing emphasis in both fields on data architecture development, data integration, and data documentation.*

*MIS and GIS development efforts can definitely benefit from a common focus on end-user systems and a blending of approaches. The Division of Forestry is currently developing and implementing plans to merge MIS and GIS staffs and approaches. The Land Management Information Center is integrating a large variety of existing tabular ground water data into a comprehensive geographic database. The goal of these efforts is to merge the best of what MIS and GIS have to offer into integrated, user-oriented systems.*

## Introduction

The origins and evolution of *management information systems* MIS) and *geographic information systems* (GIS) are quite diverse and seemingly unrelated. But, MIS and GIS shops within and across organizations have more in common than they may realize, and have much to gain by pooling their experience, expertise, and resources in pursuit of common information systems and organizational goals.

By contrasting MIS and GIS fields of endeavor in terms of the traditional definitions, categories, and histories of each, and then contrasting goals and resources, the concepts which enforce the perception that MIS and GIS are very different creatures can be brought to light. With this background as a basis, however, it becomes plausible to recognize an emerging common focus in both fields: the emphasis on *data* as an important organizational resource in its own right. And, with this realization, it becomes easy to see the advantages for MIS and GIS activities to work together toward the common good.

## Definitions, categories, histories

### MIS definitions

Management information systems are usually discussed in the context of business organizations, with examples like accounts payable, parts inventories, etc. Even though at first blush these business functions may seem far removed from the missions of natural resource organizations and other public sector agencies, they do in fact occur in all sizeable organizations (although a forest inventory may replace the manufacturing parts inventory). Most organizations do have a MIS function, whether it be an internal unit or a service bureau of some kind. The discussion below may sound anchored to the business world, but parallels to almost any workplace can be made. In fact, the typical business orientation of MIS may contribute to the perceived differences between MIS and GIS.

In an introductory MIS textbook, Murdick [1986] defines a management information system as a "system that monitors and retrieves data from the environment, captures data from transactions and operations within the firm, filters, organizes, and selects data and presents them as information to managers, and provides the means for managers to generate information as desired." This definition describes an information system that provides for ongoing transactional data entry and management-oriented information summaries. The system described here encompasses the flow of data, either in paper or electronic form, from its generation in daily business activities through its synthesis into summarized information useful to management.

James Wetherbe [1984] of the University of Minnesota describes a management information system as "a federation of information subsystems that constitutes an overall organizational information system." The subsystems he refers to are organized along both functional (i.e., finance and administration, manufacturing, marketing, research and development) and managerial (i.e., operating, middle, and top) dimensions. The emphasis here is on an interconnected system of information subsystems, which interact by receiving information from and providing information to each other.

### MIS Categories

Management information systems can be broken down into a number of categories. In **The Management of Information Systems** [Dickson & Wetherbe, 1985], the authors identify several levels of systems. *Clerical systems* are electronic incarnations of manual systems, such as transaction processing systems. They capture the daily data flow of a business. (The authors note that 95% of all computer applications today could be classified as clerical.) *Information systems* go beyond clerical systems in that they provide information (as contrasted to data) that can be used in decision-making. These systems synthesize data into useful information for managers, but leave the use of such information up to the managers. *Decision support systems* go a step further by providing both information and decision-making aids (such as "what-if" query capabilities) to help managers explore problems and arrive at decisions. Finally, *programmed systems* leave the manager out of the picture altogether by synthesizing data into information, analyzing the problem by applying a set of decision rules to the information, and producing a decision. Inventory reordering systems that automatically reorder material when the on-hand balance falls below a certain level are a simple example. The authors point out that programmed systems are primarily research topics (related to artificial intelligence and expert systems) rather than operational systems. The authors also repeat Wetherbe's [1984] theme that a management information system may be considered a federation of subsystems of the above types in an organization.

The point of this breakdown, for purposes of this paper, is that management information systems contain the elements of low-level, repeated data entry, synthesis of data into information useful for management decision-making, and, *sometimes*, decision support aids such as "what-if" capabilities.

### MIS history

Management information systems as we know them today always involve electronic computers, but their roots can be traced to traditional accounting systems (which in turn began with the invention of the double-entry bookkeeping system in 1494). This is not to say that MIS has 500 years of history behind it. Rather, MIS is the result of applying computer technology that emerged in the 1940's and '50's to in-place manual business tracking systems. Not surprisingly, many early MIS people had accounting backgrounds.

## GIS definitions

### Personnel

The type of person who has traditionally been involved with MIS development is one who is subject-area independent; who assists in the design of tools and a systematic approach for the purpose of assisting clients in meeting their needs for customer-defined information. The typical MIS expert has a background in programming and systems analysis, and is more likely to be a member of a MIS support unit rather than a practitioner of the organization's principal business.

Typically, GIS has evolved at the hands of end users who, out of necessity, have designed their own tools to solve specific problems. These subject-area experts are scattered across many disciplines and harbor many varied interests. That is one reason why today there are over 180 software systems which claim to be GIS packages listed in the most recent **International GIS Sourcebook** [GIS World, 1991]. Very few of these products share any compatibility at all. Only recently, with the emergence of a few standard packages, has the proliferation of home grown software systems slowed.

## Data as a resource

Clearly, the two disciplines have evolved from two very different directions. However, as the entire field of information system processing matures, an emerging common focus is becoming evident in both fields: the emphasis on *data* as an important organizational resource. In this view, data that describes an organization's activities and resources represents an enormous investment that must be managed and protected. Data collection and maintenance costs can represent as much as three-quarters of the budget for a typical MIS or GIS project. The view of data as a resource is exemplified by an increasing emphasis on three related system design goals: data architecture development, data integration, and data documentation.

### Data architecture

The use of entity-relationship analysis [see Everest, 1987] to develop data architectures isemerging as an important early step in system-building in both fields. In this approach, subject-area data entities (classes of objects, people, places, or events) and their interrelationships are specified by practitioners in the subject area. This analysis occurs prior to, and is independent of, the specification of application systems, software, or hardware. The goal is to produce a data-centered model of the subject area. Such a model will likely be quite stable over time, and will be a good foundation for subsequent data collection efforts and systems development. This focus on data architecture is intended to lead to the development of more robust systems, one of whose desirable attributes is data integration.

### Data integration

Historically, the computerization of data processing in many organizations has occurred by means of independent MIS projects undertaken at different times in different functional areas. When looked at from an organizational perspective, this has resulted in fragmented systems, one of whose major flaws is redundant and sometimes inconsistent data entry and storage. One of the benefits of a data architecture approach, if conducted at the appropriate levels in an organization, is that it will lead to non-redundant data collection and storage organized around entities, rather than around particular functional applications.

Similarly, GIS projects have often tended to focus on the problem at hand, resulting in data sets that may be of little use beyond the current project. One approach to integration of geographic data involves the design of data sets and data collection efforts for multiple use. Entity-relationship analysis is useful here. A study of a subject area's data entities and interrelationships is likely to lead to wider participation of practitioners in related subject areas. This in turn can result in coordinated data collection efforts that should cost less overall than would separate efforts. One example of such an approach in the development of a spatial database is the ongoing effort at the U.S. Geological Survey to design the Digital Line Graph-Enhanced (DLG-E) data structure [Domaratz and others, 1990].

A second approach to geographic data integration is the inclusion of geo-referencing data such that multiple data layers from different projects can be overlaid. It is obviously impossible to involve *all* potentially interested parties in every GIS project. However, registration to common control points and the use of common coordinate systems across projects will provide future overlay capability for data sets that are designed separately. An example of cross-project coordination is occurring in Minnesota's Land Management Information Center (LMIC), which is developing a role as a clearinghouse for geographic data bases within the state. LMIC's efforts to make GIS professionals aware of existing and planned data sets, and to promote the use of common control points and coordinate systems, will greatly increase the potential for integration and wide-range use of Minnesota's geographic databases.

### Data documentation

If an organization's spatial and non-spatial data are considered to be an important resource, then data documentation is a necessary tool for managing and protecting that resource. Entity-relationship analysis should be carried through to the creation and population of a *data dictionary*, one which stores metadata (data about the data that an organization collects). This will document the specification of data entities, the attributes used to describe them, and the relationships among them. Such a dictionary will greatly reduce the tendency to "reinvent the wheel" when new systems or projects are undertaken, and will promote data integration.

A system and data set catalog is a data documentation tool useful at a more aggregate conceptual level than the data dictionary. While a data dictionary describes data at the row and column level, a system catalog describes data at the table and database level. Such a catalog describes existing systems and data sets, the types of information that they contain, the sources (lineage) of the data, and the formats in which data are available from the systems. This type of metadata will help system and project developers learn about potentially related previous efforts, and secondarily can help end users discover where to find needed information.

## Working together toward integrated systems

The backgrounds, goals, and resources of MIS and GIS do tend to be quite different. Sometimes, it seems as though the two disciplines are speaking different languages. However, shrinking budgets and expanding information needs dictate that organizations must strive to pool their resources and approaches to develop systems and projects that will provide them with timely, well-integrated, and cost-effective information. MIS and GIS units, within and among organizations, can work together effectively by, 1) developing a common focus on end-user systems and 2) integrating their approaches.

### Develop common focus on end-user systems

MIS professionals see a system as a set of turnkey, end-user applications. GIS professionals often define a system as a set of spatial analysis tools. Part of the difference stems from the fact that the users tend to be different in these fields. MIS users are usually line member of the organization who are not computer literate (as opposed to staff MIS developers). GIS users and developers are often the same people: practitioners who have learned to use the spatial analysis tools to solve their own problems. From this perspective, a commercial GIS software package such as ARC/INFO *is* a set of turnkey, end-user applications. They are definitely more flexible than typical MIS applications. The trade-off is the time investment needed to master the tool set. For those subject-area experts who have already made that investment, the project-oriented, ad hoc analysis type of GIS is still a viable information management tool.

A similar argument could be made in the MIS world that the new crop of powerful databasemanagement systems, coupled with fourth generation procedural languages (4GLs), constitute "systems" in the hands of computer-literate end users. However, once again, the trade-off for flexible, hands-on access to data is learning the tool set.

Middle- and top-level line managers do not have the time to become power users of sophisticated spatial analysis or database management tools. Organizations do not have the resources to make high-level computer literacy a requirement for access to organizational information. The technical components of formal GIS and MIS systems should be developed by information system specialists, whose primary responsibilities include the mastery and use of sophisticated development tools. However, it must be remembered that those tools are only the means to an end. The main focus of those systems must be the clear presentation of useful management information to end users. These users must be intimately involved in the development process. They should fully participate in development projects to insure that their information needs are met.

### Integrate approaches

Given that MIS and GIS professionals agree to a common focus on end-user systems, the next step in working together is an integration of traditional MIS and GIS approaches to systems development. MIS methodology can contribute the formal approach to development embodied in system life cycle analysis. GIS methodology can temper the often-perceived ponderousness of this approach with a contribution of the technique of prototyping. This technique can help the user members of a project team to refine quickly their requirements for screen and report layouts and process flow. These requirements, once refined, can be included in the life cycle analysis.

MIS professionals can also contribute their expertise in automating and "bullet-proofing" repeatable functions in a system, so that they can be performed without confusion and without error by non-computer-literate users. GIS professionals can contribute their expertise in developing ad-hoc analysis and reporting tools that will give users the flexibility to explore information in new ways.

## Making integration work

A brief discussion of two ongoing system development efforts in Minnesota will serve to illustrate how the approaches to integration discussed in this paper are being implemented.

### DNR forestry

The Minnesota Department of Natural Resources, Division of Forestry has been developing and implementing plans to integrate their MIS and GIS functions over the last three years.This effort began as part of a strategic planning project for information systems in the Division known as the Forestry Information Systems Blueprint [Minnesota DNR Forestry, 1989]. The MIS and GIS units in the Division developed separately and worked independently for several years. The MIS unit was staffed by traditional systems analysts and programmers. The GIS unit was staffed by line forest inventory personnel eager to use a new tool to maintain an operational forest inventory. The Blueprint initiated efforts to combine these two functions along several fronts:

- Merge MIS and GIS staffs into a single FIS Forestry Information Systems) unit.
- Cross-train MIS and GIS practitioners so that they can share tools and approaches.
- Merge separate MIS and GIS hardware (two brands of minicomputers) into a single platform (a RISC-based UNIX workstation) to reduce maintenance overhead and provide a common operating system environment for development.

- Merge separate tabular data handling tools (COBOL and INFO) to a single yet-to-be-determined relational database management system (RDBMS) in combination with a fourth generation procedural language, in order to provide a common set of powerful development tools for handling non-spatial data. Use of the RDBMS will facilitate implementation of entity-relationship data models. Use of the 4GL will facilitate prototyping.
- Given the infrastructure in the above points, design and develop new information systems (and redesign existing ones from the ground up) to include spatial and non-spatial data integrated so as to produce meaningful management infor-mation to end users. MIS and GIS development staff, as well as end users, will work together on project design teams. They will use life cycle analy-sis in conjunction with prototyping for standardized, yet flexible, design methodology. They will use the data architecture approach to define spatial and non-spatial entities, attributes, and relationships. They will build bullet-proof, repeatable procedures ***and*** ad-hoc query and reporting capabilities into their systems. Above all, they will focus together on the goal of provid-ing useful information to help end users perform their jobs.

**Ground water clearinghouse**

A number of state agencies in Minnesota have responsibilities for protecting the quality and quantity of the state's ground water resources. To carry out their responsibilities, these agencies have collected a variety of ground water data in a variety of computer compatible formats. Unquestionably, these collections of often disparate and incompatible data have been useful in carrying out the immediate duties of each managing agency. However, they prove rather inaccessible to potential users who wish to merge data sets for purposes other than originally designed.

Minnesota's *Ground Water Protection Act of 1989* mandates that water quality monitoring data be made more widely available by integrating them into a comprehensive geographic database plan. LMIC, with the cooperation of those data-generating agencies, is currently integrating the most important of these databases in a project known as the *Ground Water Clearinghouse*. The integration of some ten major state databases and an additional 30 to 40 smaller data collections is focusing on the following goals:

- Make important ground water data collections more readily accessible through a single access point.
- Merge small, disparate water quality data collections.
- Improve coordination among agencies on database development.
- Provide an index to the content of other related databases.
- Improve ground water monitoring coordination.
- Integrate tabular water data into a GIS to demonstrate a graphic display and map generating capability.

The results of this effort will be an information system: one which truly integrates a host of diparate databases, providing to a wide rand of data users on-line access to a valuable state resource. By merging tabular data with corresponding spatial information and providing access to commercial GIS software, this system will provide the added benefit of being able to display the contents of the data base and will provide users with easy and flexible access to view the information contained within. This type of database-wide viewing capability is a feature which has not been available in the past.

The development of an information system composed of a comprehensive, well-designed database, offering a structured, yet robust set of query and graphic display options in an easily accessible, user-friendly environment will act as an important illustration of the potential power of merging the best of what MIS and GIS have to offer.

## References

Burrough, P.A. 1986, Principles of Geographic Information Systems for Land Resources Assessment. Oxford University Press, New York, NY.

Dickson, G.W., and J.C. Wetherbe, 1985, The Management of Information Systems. McGraw-Hill Book Company, New York, NY.

Domaratz, M.S., D.E. Hair, G.B. Chappell, C.J. Cialek, and K.C. Neff, 1990, DLG-E Internal Exchange Structure and Format. U.S. Geological Survey internal document.

EOSAT. 1991. "EOSAT and Terra-Mar to Study GIS Economics". Earth Observation Satellite (EOSAT) Company News Release 91-12. Lanham, MD.

Everest, G.C. 1987. "Users Do Logical Database Design", Proceedings, CAD/CAM Database '87 Conference, San Francisco, C.A.

GIS World, 1991. 1991-92 International GIS Sourcebook, GIS World, Inc. Fort Collins, CO.

Minnesota Department of Natural Resources, Division of Forestry, 1989. Forestry Information Systems Blueprint. St. Paul, MN.

Murdick, R.G., and J.C. Munson, 1986 MIS Concepts and Design. 2nd ed. Prentice-Hall, Englewood Cliffs, NJ.

Wetherbe, J.C. 1984, Systems Analysis and Design: Traditional, Structured, and Advanced Concepts and Techniques. 2nd ed. West Publishing Company, St. Paul, MN.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Trends in GIS technology. What are the impediments?

**Carl Reed, Ph.D.**

President
Genasys II, Inc.
2629 Redwing Road
Suite 330, Fort Collins
CO, USA 80526

### Abstract

*Dozens of articles have been written and presented on the Future of Geographic Information Systems Technology. The majority of these papers describe the impact of faster, less expensive machines; new applications; integration with modeling; new algorithms; and so forth. Some of the more recent articles discuss such topics as open systems, standards, and distributed databases. A few even go so far as to proclaim that GIS will "save the world". The common thread in all cases is that GIS will get faster, bigger, and better, and that more people will have access to the technology. This paper, while touching on actual trends, will focus on impediments to the future of GIS technology. Key issues such as data sharing, the lack of trained GIS professionals, fear of change, and competition will be addressed.*

## Introduction

A GIS vendor has been recently proclaiming that GIS will be available to the masses and that GIS will solve many of the world's problems. A nice thought, but we are a long way from realizing such a utopic vision for GIS. In the "real" world, economics, politics, and education will have a much greater impact on solving the world's problems. GIS is only a tool that can be used to help make more informed decisions, albeit, a very powerful tool. Even so, there are a number of impediments to the future use of GIS on a broad scale. These impediments are not purely technical but also relate to current conceptions (and misconceptions) about what GIS is, what GIS can and cannot do, and how GIS use will evolve into the future. This paper describes several major impediments to the future growth and use of GIS and suggests some solutions. The paper concludes with the more traditional forecast of evolutions in GIS technology.

## Data sharing

### The problem

A commonly held belief is that the most costly aspect of implementing GIS technology is the collection and conversion of data. (Actually, over the long term the most costly item is staff, but more on that later.) For a large county or city, the cost can easily be in the millions. As such, data becomes an extremely valuable resource. Many organizations are looking to recover some of this cost. At the same time, many new users of GIS technology want to use existing digital map data. For the new user, this decreases implementation time and implementation cost. In the case of well-established GIS groups, there is an increasing desire to share data. Overall, the trend is toward an ever-increasing requirement to share data.

It would be a perfect world if data sharing were seamless. This is not the case. A major impediment to the future growth and use of GIS is the current inability to easily share data, even within the same organization. Because it is not easy to share data, there is a very nasty tendency in the industry to force new GIS users to use/buy a particular GIS solution. The justification is, "You can use our data immediately." For example, many state agencies have been collecting digital map data. These same agencies will cost share GIS implementation with a county. The caveat is, "Oh, by the way. You must buy Vendor X's GIS so that we can easily share data." This attitude not only limits competition, but also misrepresents the supposed simplicity of data sharing in a common software environment. For example, states and counties tend to work at different levels of resolution, require different classification schemes, and have different applications.

### The solution

Fortunately for both GIS users and vendors, there are definite trends in GIS technology that make data sharing increasingly easier to accomplish. These are binary translators, emerging Federal standards, and an increasing awareness that "meta-data" models are perhaps the ultimate solution for data sharing.

Binary translators: Binary translators allow users and/or software vendors to program software modules that read the internal binary data structure in one system and write directly into the binary data structure of another system. These translators tend to be very easy to use and process data much more effeciently. Obviously, to be able to have binary translators requires that vendors either publish their proprietary data structure format(s) or provide the tools for direct access to the binary structure.

Some vendors, such as Intergraph, feel that providing this information and access is important to the mapping and GIS industries. As such, a number of DGN translators exist for sharing data with other vendors' GISs. Unfortunately, some vendors feel that their data structures are proprietary and will not release any documentation on those structures. Perhaps the real reason is that these vendors wish to maintain a "lock" on the market by enforcing data sharing in proprietary binary formats to maintain a competitive edge. If this is true, such an attitude is deleterious to the long-term-growth of GIS. It inhibits competition, limits potential, and, most importantly, sidesteps the real issues of data sharing.

I suggest that all GIS vendors either publish their data structures or provide a set of inexpensive programmer interface tools for building binary translators. Such information or tools will also promote the concept of meta data (discussed below).

SDTS: Another potential solution for data sharing, especially in the use of Federal data, is SDTS (Spatial Data Transfer Standard). This standard, promoted by the Federal government, theoretically has slots not only for the map topology but for symbology and attributes as well. These latter two areas have been major shortcomings in previous generalized translators. To date, SDTS has been through various committees as well as ANSI review. The net result is that SDTS shortly will be both a FIPS and an ANSI standard. As such, any organization or vendor faced with data sharing will have to write bi-directional SDTS translators.

The downside of the SDTS is that it uses ASCII files for the actual interchange. Therefore, translation will be quite a bit slower than a binary translator. It also results in more processing steps and can require large amounts of disc space for file storage. In addition, because SDTS attempts to be the translator that accommodates the many differences and variations in spatial data, it is problematic whether many groups will actually implement the full specification.

Meta data: Meta data is the next step beyond binary translators. In a true meta-data environment, no data is actually translated from one file format to another. The data remains in its original format on disc. For example, suppose an organization has an Intergraph system and a Genasys system. In the current mode, a DGN binary translator is used to move data from one system to another. In a meta-data environment, Genamap would read data directly from the Intergraph database, and vice/versa, on the fly.

The implementation of the meta-data approach draws on what is becoming the standard mode of both graphics interfacing and RDBMS interfacing: a client/server model. In our example above, a server would handle the interfacing of one GIS to another. This server would receive a request for data, read the appropriate data in binary form from the target database, and send it back to the client. The final step, sending data back to the client, also involves putting the data in a structure the client understands.

This approach has allowed organizations to use multiple RDBMSs to maintain their enterprise databases. If the same philosophy is used in GIS, an organization can use the best tools for each application area while easily sharing data with minimal overhead and aggravation. It would be possible for an organization to have multiple CAD/CAM systems and multiple GIS solutions! Further, if one looks beyond a single organization and considers a multiparticipant environment, such as cities within a county, the benefits are even greater. Each entity could purchase the GIS and/or CAD solution it felt was best suited for its environment without having to worry about data sharing. As long as each group was a node on a network, the meta-data approach could be implemented.

Obviously, for this approach to work, the vendors (once again) must be willing to provide the documentation or tools to allow the implementation of a meta-data environment. There are other impediments to implementing a meta-data environment, especially in a multi-participant situation. One is the legal issues surrounding the use of someone else's data. The legal issues focus on accountability and liability. Another is politics: will different organizations be willing to share data in such a manner? The institutional issues would have to be worked out on a case-by-case basis.

## Decreasing fear of change

This is not the fear of adopting GIS technology. It is the fear of changing to another GIS solution or augmenting the existing system with another GIS after using a particular system for years. Some vendors play on this fear by talking about massive retraining or massive problems in data conversion. This may be true in some situations, but in the majority it is not. With the advent of binary data translators, many of the data conversion issues disappear. If a meta-data approach is used, they disappear almost completely. With the advent of easier-to-use user interfaces, the retraining issue is also greatly reduced. This is especially true if one considers that the majority of training is in concepts and procedures. Conceptually, polygon overlay, buffer analysis, or distance searching are almost identical between systems.

In 1982, I worked with several Alaska state agencies. These agencies worked with multiple GIS solutions. The reason was simple: they used each system for what it was best at. It did not bother them in the slightest that they required different vendor systems. The same is true today - no GIS does it all! There is no reason, especially with binary translators or with the potential for meta-data environments, that an organization cannot switch GIS solutions or augment its existing system with another vendor's GIS.

Unless organizations consider this model, they not only will be limiting their own future growth but that of the GIS industry in general.

## Revenue generation

One of the "hot" topics in GIS right now is revenue generation. This trend will definitely continue into the future. This is because funding of GIS technology on a year-to-year basis is an ongoing task. One of the major contributers to GIS failures is lack of funding. Therefore, many organizations are looking at revenue generation as a long term solution to funding ongoing GIS activities.

The concept is that an organization, such as a county, has invested considerable time, money, and effort to build a digital map database. They would like to recoup this original investment as well as cover the costs of maintaining the database. Approaches for revenue generation such as front counter applications and the selling of data to the public are usually discussed. On the surface, revenue generation appears to be a funding panacea. However, there are some major blocks.

Resistance: Typically, the first impediment to revenue generation is, "These data were collected using the tax payers' money. I am a taxpayer. Therefore, why are you charging me for something I already paid for?" This impediment can be handled by mandate, by legislation, and/or by adding value. Basically, the taxpayers have to be educated to the fact that it costs money to maintain this database that is used to improve services to the taxpayer. Further, as in the case of the front counter application, additional value is being added - data are being integrated from many sources into a product. This product is being delivered much more quickly and with much greater accuracy than in the past. Therefore, there is a fee for this.

Copyright: Most open records laws provide for unlimited access to government data. There has traditionally been quite a bit of confusion about how copyright law and the open records laws work. There is an excellent discussion by Ms. Lori Dando on GIS, open records, and copyright law in the 1991 Urisa Proceedings (Volume 4, page 1).

Liability: The threat of a lawsuit is ever present in GIS, especially when the data are being used for decisions related to zoning, landuse, assessment, and development. When data are provided to the public and there is a fee, steps must be taken to minimize the risk of a lawsuit. This is usually done through a disclaimer on any generated product. Unfortuately, no GIS currently tracts error propagation or maintains internal information on errors in the data. This whole concept of error analysis in GIS and automated mapping has been ignored in available systems. If this functionality were available, then accurate statements regarding data accuracy could be made for any generated product. Doing so would greatly minimize liability and greatly improve the chances for revenue generation.

## Lack of trained GIS professionals

Even though more and more universities and colleges have developed excellent programs in GIS technology, the GIS market place has grown so rapidly that the gap between the availability of trained GIS professionals and the need for these professionals has been ever increasing. This gap is on all fronts. Vendors require software engineers with GIS knowledge. User organizations and vendors both need application specialists. User organizations require geographic database management specialists. Vendors require staff for GIS training and support. Finding trained GIS personnel with the correct experience background is very difficult. As a case in point, I recently met a GIS "professional" whose academic training was in American Literature!

For any GIS user environment to succeed, experienced staff is necessary. These staff should not only understand GIS but, even more importantly, they need to understand the application area they are working in. Unfortunately, in many cases this is not true. Typically, an organization will assign existing staff to become GIS specialists. These individuals know their application extremely well, but know nothing about GIS. The converse is when the organization hires a new staff member. This new staff member may be a trained GIS specialist but probably knows little about the application area he/she is to work in. In either case, the result is a large amount of on-the-job training. With budgets getting tighter and tighter, such on the job training may become a luxury. Regardless, the end result is a much more time consuming and frustrating GIS implementation than either the vendor or the customer expected or wanted.

The current solution to the problem of a lack of trained GIS professionals is that vendors hire from their installed based, which hurts the installed base, and GIS user organizations hire from each other. Hopefully, as more trained GIS specialists graduate from the new GIS programs, this situation will change. However, even if the universities could supply the majority of the trained specialists required to support the growth of the GIS industry, there are still some major problems.

One problem is that the majority of the students graduating with GIS experience do not have application experience. While environmental issues are very relevant and interesting, the majority of GIS technology is purchased by local government, utilities, and the Federal government. Too many students have no experience in these areas. Thus, when they do enter the job market, they are not adequately prepared. One solution is for students to enter intern programs that will allow them to develop application skills in an industry area.

Another problem is that GIS programs tend to be affilliated with one department within the university. Therefore, the GIS program tends to be slanted toward the philosophy, politics, and graduation requirements of that department. This creates a problem because in reality, GIS is a multi-disciplinary field. Courses in geography, mathematics, civil engineering, business administration, and computer science are all equally relevent. Unless the GIS program supports and promotes multi-disciplinary study and accreditation toward a degree, then the GIS program will be diluted and the GIS student will not be developing the tools required in the market place.

The solution is that the GIS department works with the student to develop a program suited to that student's desire to gain employment in a certain market sector. In this way, the student will gain knowledge and experience not only in GIS but in a specific application area. Further, these application areas should be focused in the major GIS markets - local government, utility, and commercial sectors - and not marginal application areas.

## Business ethics and standards

The GIS market place is becoming increasing competitive. As such there is more and more pressure on the vendors to maintain market share, much less increase market share. The result of this pressure is that there appears to be an increase in predatory pricing and product dumping. Predatory pricing is the practice of intentionally underbidding the competition, even at a loss, to win business. Product dumping is closely related and refers to the practice of selling product at a price less than the cost of manufacturing and maintaining the product.

Both are techniques that could be used by small companies to try to obtain market share and by the larger, existing companies to maintain or increase market share. An example of the first case is a new vendor trying to penetrate the market and obtain an initial market share. An example of the latter case is a large, established vendor trying to prevent another vendor from obtaining business. In either case, I feel the business practices are in many cases unethical and in some cases illegal.

The cornerstone of the free enterprise system is competition. Without competition, vendors will not provide the best product or the best service - there is no reason to. Without competition, the end user suffers by only having a limited selection from which to choose a solution. The user further suffers because the limited selection will not be as good and will not be supported as well as in a truly competitive environment.

The topic of standards and ethics (especially the ethics) is one that the GIS industry seems to want to ignore. It cannot be ignored. A strong GIS industry needs a strong, healthy competitive environment. Many industries have agreed upon standards of code/ethics as well as an oversight organization to resolve conflicts. Perhaps the GIS industry should do the same.

## Conclusion

GIS is a robust, rapidly growing industry. As with any maturing industry, it is beset with a number of issues and problems that need to be resolved in the next several years if it is to continue growing. A number of problem areas have been identified. All of them have solutions. Indeed, there are a number of initiatives already progressing within the GIS industry that directly impact data sharing (SDTS) and GIS education (the National Centers of Excellence). Let us hope that these initiatives continue and work together as an industry to resolve other impediments to the growth of GIS technology.

# D2 Evolving methodologies

**The use of stream drainage patterns in a GIS program package for the delineation of geologic faults and fractures**

Richard Thiessen et al., Washington State University, Pullman, WA, USA

**A decision support system using neural nets and genetic algorithms for multi-objective State Park management**

Randy Gimblett et al., Ball State University, Muncie, IN, USA

**Parser design strategies for GIS natural language interfaces**

William Lydiatt, Interface S.A., Santiago, Chile

**Phase II EPPL shell: A GIS application for the field resource manager**

Dave Martodam, Minnesota DNR Forestry, St. Paul, MN, USA

D2 Evolving methodologies

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# The use of stream drainage pattterns in a GIS program package for the delineation of geologic faults and fractures

**Richard L. Thiessen** (presenting author)

**Eric R. Rieken**

**Douglas S. Neves**

Geology Department
Washington State University
Pullman, WA 99164-2812, USA
Tel: 509-335-4947, 335-8850, 335-3009
Fax: 509-335-7816

## Abstract

*The delineation of geologic faults and fractures is critical for a number of environmental and natural resource management applications. WSU, in conjunction with Pacific Northwest Laboratory, is generating a three-dimensional (3D) GIS tool, Geologic Spatial Analysis (GSA), designed to automatically identify geologic fractures using a wide variety of data bases, including but not limited to digital elevation models (DEMs), remote sensing images, gravity and magnetic data, subsurface geologic information, and earthquake foci. Fracture controlled drainage networks are detected using an expandable kernel applied to DEMs. Smaller detectors (7 pixels across) preferentially find vectors delineating narrow upland canyons. These can be combined with major trunk streams found with wide detectors (71 pixels). Advanced edge detection routines are applied to DEMs, images, and geophysical data bases in order to delineate possible fault controlled gradients and other features. These can be registered to DEMS to generate 3D vectors. A counting cell system examines the spatial relationships between the data bases in order to find three-dimensional planar alignments of earthquake foci (point data) in combination with the potential fractures (vectors) detected in the other data bases. Surface traces of fracture systems as well as 3D models are produced. The result is a computer automated 3D delineation of the location and orientation of major geologic faults and fracture systems in a study area. These can be directly utilized in resource management and environmental applications.*

## Introduction

### Geologic spatial analysis

GIS systems are being developed and applied to a wide range of disciplines. Washington State University (WSU), in cooperation Pacific Northwest Laboratory (PNL) is developing a major GIS tool targeted at geological and environmental applications. The Geologic Spatial Analysis (GSA) research project is focused on conducting comprehensive automated three dimensional (3D) analyses of regions using a variety of geologic data bases and formats that can be referenced by location and elevation/depth. These include digital elevation models (DEMs), seismic foci, remote sensing images, geophysical data, seismic reflection profiles, field studies, and borehole fracture intersections. Based on these data, the GSA techniques examines the spatial relationships in order to determine the location and orientation of geologic fault and fracture systems in 3D space in order to develop a regional geologic model for the user. The GSA project provides researchers with GIS tools that are computer automated and combine a wide variety of data types. GSA determines the 3D geometries of probable geologic faults, which are essential for management of energy, mineral and groundwater resources. They are also crucial for the assessment of environmental impacts and seismic risk, as well as basic geologic research and mapping.

Basic research being conducted on GSA techniques at WSU has been sponsored in the past by the U.S. Department of Energy's Office of Basic Energy Sciences (OBES) (grant DEFGO6-87ER13796 for $107,177 which supported 30% of total research). The research project is closely coordinated with a companion program at Pacific Northwest Laboratory (PNL), which has sponsored several specific GSA projects at WSU (BN9191-A6, $15,000; 009191AI1, $1329; 097718-A-L2, $1997). PNL has also funded a number of faculty and graduate student NORCUS (Northwest College and University Association for Science) fellowships, including Richard Thiessen and Eric Rieken.

A number of GSA techniques are past the basic research phase. These include COPLANE (identification of coplanar fracture controlled valleys), SEISPLN (seismic foci alignments), and LINEAMENT (analysis of manually delineated lineaments in 2D and 3D), as discussed below.

Erosion is often localized by fault and fracture systems producing linear stream valleys. If a single fault surface controls a number of valleys, then they should be coplanar in 3D space. Our COPLANE analysis sequence examines this relationship. This initial GSA concept (Eliason, 1984; Eliason and Eliason, 1985; Eliason and Foote, 1986; Eliason and Thiessen, 1987) was devised by Jay Eliason (now at Geologic Analysis and Consulting Services, Deary, Idaho) as his WSU thesis. Operational GSA programs have been generated to detect low points in digital elevation models (DEMs). These lows are connected into stream valley strings which are used to define 3D vectors. These vectors are examined for coplanar alignments using the cross product routines of our COPLANE program package. Automatic clustering routines look for groupings of coplanar alignments, indicative of major fault systems (Foley and others, 1988; Thiessen, 1991). A current research topic is a variation of the COPLANE sequence which will identify range front faults, such as the Wasatch Fault in Utah. COPLANE has been utilized in a variety of modes at a number of study sites (Eliason and Eliason, 1985; Eliason and Thiessen, 1987; Thiessen and others, 1988, 1989A, 1989B, 1989C, 1989D, 1989E; Foley and others, 1988; Beaver and others, 1989).

Hand picked lineaments can be digitized and analyzed using our LINEAMENT package (Thiessen, 1986; Johnson, 1988). Mapped faults and geophysical linears can also be studied. GSA algorithms delineate the high confidence lineaments observed a large number of times on multiple images. Lineament analyses of a number of sites (Thiessen and others, 1985, 1987, 1988, 1989C; Pitz and Thiessen, 1986), including multiple interpreter, multiple image studies of western Washington (Rieken and Thiessen, 1990; Thiessen and Rieken, 1991) have been conducted. The lineaments can be registered to topography in order to create the 3D vectors required by COPLANE, as we have done for an analysis of subsurface faults and surface lineaments in one study (Thiessen and others, 1989D).

One of our automated set of computer programs, SEISPLN, has been developed for the identification of 3D alignments of point data, particularly seismic foci (Rieken, 1985; Thiessen and Rieken, 1991). This package will work for any number of planar or sub-planar structures. The SEISPLN technique is based on rotating a 3D elongate grid system through the seismic foci. The number of events in each cell is tabulated, and those cells which contain statistically significant numbers of events are considered to be potential fault surfaces. A multiple linear regression is applied to the set of foci in each delineated cell to model the 3D geometry of each of the inferred fault surfaces. Application of this technique to the subducting Juan de Fuca (JDF) plate beneath western Washington is providing us with a more accurate model (Rieken and Thiessen, 1990, 1991, in prep; Rieken and others, 1991) of the Benioff zone on which to base our crustal fracture models.

Virtually all GIS applications and systems have to deal with spatial analysis of different data formats (raster, vector, point), and the GSA packages are no exception. SEISPLN looks for three-dimensional alignments in point data, and vectors are derived from raster data and then analyzed with COPLANE and LINEAMENT. In the past, the results of the individual methods have been visually compared. Current research involves the integration of the SEISPLN technique with the analysis of topography and lineaments in a true GIS multi-data set format. The end points of the vectors defined from valley bottoms have been used as individual points in combination with earthquake locations in order to better define structures. In the future, vectors will be input directly into SEISPLN by directly determining which counting cells they could lie within. This will be a three-dimensional GIS tool.

The active faults responsible for the crustal earthquakes of the Puget Sound and southwest Washington will be the application study for this project. Input point data were earthquake foci collected by the Washington Regional Seismograph Network (Steve Malone, pers. comm., 1990). Vector data includes valley bottoms detected on 1:250 000 scale DEMs and high confidence lineaments derived from an eight image, 50 interpretations lineament study already conducted by us (Rieken and Thiessen, 1990; Thiessen and Rieken, 1991). The result is a clarification of one of the crustal faults, the Nisqually Valley structure, in this critical region using GSA techniques.

## Drainage networks

Stream drainage networks are an integral part of many GIS applications, and a number of methods have been devised to automatically define their valleys based on DEM data. Because stream valleys are often controlled by geologic fault and fracture systems, valley segments are an integral part of WSU's GSA analysis system.

The technique used by most studies (O'Callaghan and Mark, 1984; Jenson and Domingue, 1988; Morris and Heerdegen, 1988; Lammers and Band, 1990; Tribe, 1991) is based on conceptually pouring water onto the DEM and see where it concentrates. First, the drainage direction for each grid point in the DEM is determined by looking at which of the eight nearest neighbors (in a 3 by 3 kernel) is lowest and is also lower than the original grid point. This is then used to calculate the total number of cells that drain into the grid point including drainage through intermediary cells. Locations that are above some threshold value are assumed to be stream drainages and can then be thinned to a single pixel thickness and connected into stream segments. Many DEMs are noisy, with pits and stripes having been created by the data generation process. This method has problems with these and natural drainage basins, but can be modified to step over these sinks (Morris and Heerdegen, 1988; Jenson and Domingue, 1988; Tribe, 1991). Determination of the drainage direction can be ambiguous in areas of flat topography, but one can expand the detector size (Morris and Heerdegen, 1988) until the direction becomes better defined.

Palacios-Velez and Cuevas-Renaud (1986) traced the forward and reverse flow path between two grid points. The uphill path was defined as going up the steepest slope. If the downhill path and uphill one were identical, then the path is a pure streamline. Otherwise, it could be a valley.

Haralick (1983) and Smith and others (1990) fit a higher order surface to a local portion of the DEM. This is then used to calculate first and second derivatives. Valley bottoms should be zones of rapid upward curvature, and so will have large positive second derivatives, with the maximum curvature being perpendicular to the valley trend. The valley side walls will slope in opposite directions and so the first derivative will go through a zero crossing when measured perpendicular to the valley. If both conditions are met, then the location is flagged as a valley segment and retained for further analysis.

Peucker and Douglas (1975) and Tribe (1991) passed a 2 by 2 kernel over the topography and simply flagged the highest grid point lying in the box. After the kernel passes over the entire data base, the unflagged low points are assumed to be the drainage points. This method is too simple and leads to non-continuous valley segments and inconsistent results (Tribe, 1991).

Jenson (1985), Qian and others (1990), and Tribe (1991) superimposed a 3 by 3 kernel onto the DEM. Four different directions were tested (north-south, east-west, northeast, and northwest). If for any of the directions, the center point was lower than both ends, then the grid point was flagged as a valley low point. Qian and others (1990) had a quasi-expandable detector where the average value of a block of points would be used instead of just a single value. They still used a 3 by 3 group of data. Desachy and Lammali (1989) used the same four directions, but looked at entire topographic profiles by extending their test to the edge of the DEM data base. Low points on each profile were flagged and then tested to see if they were simply local lows or true bottoms of valleys.

Most of the above methods do not yield single pixel wide valley bottoms, but instead create elongate zones of grid points. These must be thinned to single thickness and then connected into valley segments (Greenlee, 1987). WSU's GSA system operates with only the location of these valley segments. However, for most other GIS systems, what is really desired is the interconnection of the entire drainage network. A number of studies have devised techniques to topologically link the individual valleys into a stream network (O'Callaghan and Mark, 1984; Band, 1986; Morris and Heerdegen, 1988; Jenson and Domingue, 1988; Van Pelt and others, 1989; Lammers and Band, 1990; Smith and others, 1990; Qian and others, 1990). One study extended this by using an expert system to classify drainage system types (Hadipriono and others, 1990).

## Expandable detector for location of valley bottoms

One of the major problems with stream drainages is that they are made up of valleys of a wide variety of scales and sizes. Major trunk valleys will often have relatively flat bottoms that can present special problems to valley detection routines. In order to address this problem, we have developed a valley detection routine which includes an expandable kernel. The original detector (Eliason, 1984; Eliason and Eliason, 1985; Eliason and Foote, 1986; Eliason and Thiessen, 1987) was a 7 by 7 box kernel that was run over the DEM. Sixteen independent directions were tested. The central grid point was compared to both end points of each direction to see if the center was significantly lower than the two ends, and therefore possibly a valley bottom. A user specified threshold controlled how much lower the central point must be. The algorithm returns the number of directions that a valley low point is detected (out of the sixteen possible). The central portion of well formed straight valleys may be detected up to twelve times, whereas more complex ones would be detected fewer times. One therefore has a measure of whether the valley has been eroded along a planar geologic structure such as a fault, fracture, or bedding plane.

WSU has recently written a new computer program package, NODES, which allows for variation in the size of the valley detection kernel. The user specifies the diameter (in pixels) of the desired detector as well as how thick it will be. In order to define a central point, the kernel size must be odd (3, 5, 7, ... 99). The computer then generates a circle with the desired size and thickness. Any grid point falling within the circle is considered a kernel testing point and is given a value of 1 in the kernel (Figure 1). Points outside the circle are assigned a value of 0. The circle can be easily defined for grids where the north-south spacing of the grid points does not equal the east-west size. Each kernel point will have a symmetric point on the opposite side of the circle. The computer then takes this kernel and superimposes it on each successive grid point in the DEM data base. Any DEM point that falls under a kernel point with a value of 1 is tested against its symmetric point as well as the DEM point in the center of the kernel. If the center point is topographically lower than both end points by some value in excess of the user defined threshold, then the number of detections for that center point is incremented by one. Differences in distances from the center point to the kernel points is scaled so that more distant points are tested against a somewhat smaller threshold value. The maximum number of possible detections for extremely large kernels may be several hundred. In other words, the valley has been checked to see if it is a central low point for hundreds of directions. The actual number of detections at each DEM grid point can be plotted as a color image, and pixels with values within a user specified range can be analyzed. This obviously gives a powerful new GIS tool for examining the many complexities of drainage systems.

The WSU detector will reduce to the 3 by 3 kernel used by Jenson (1985), Qian and others (1990), and Tribe (1991) if one uses a detector size of 3 pixels, a thickness of 1 pixel, a threshold elevation difference of 0 meters, and the required number of detections as 1. It also reduces to WSU's original 7 by 7 detector for a kernel size of 7 and thickness of 0.85 pixels.

The valley bottom points that are detected can be thinned using several different algorithms in the NODES program. The preferential method is to retain the grid point that has the maximum number of detections, as this will be in the center of the valley. They are then linked together into valley segments for ease of storage and processing. The valley segments are vectors in 3D space, and can be utilized as input into the COPLANE or SEISPLN analysis packages.

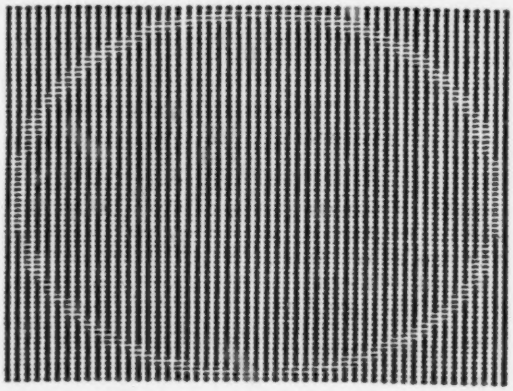
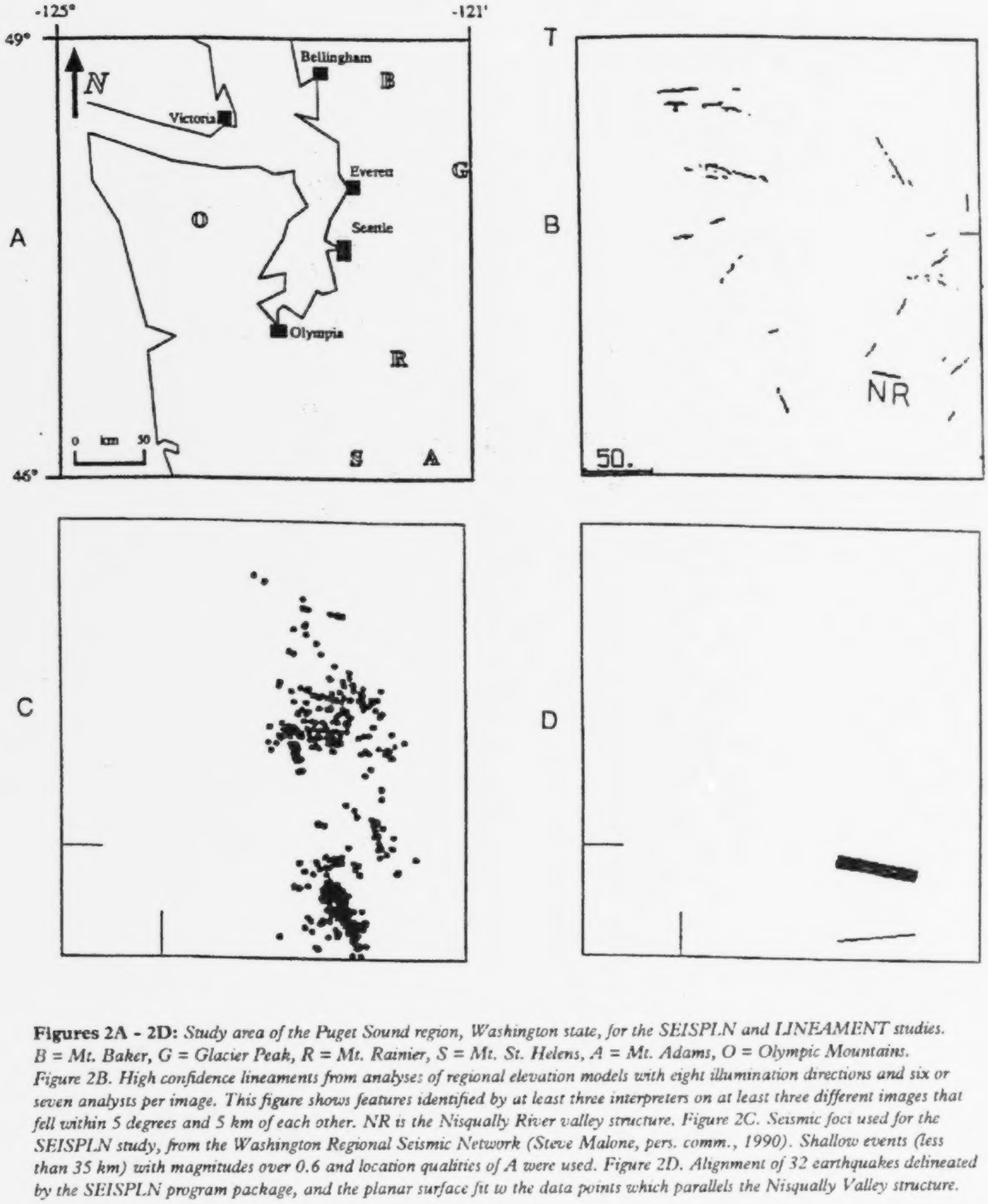


**Figure 1.** *Example of an expandable valleys detection kernel. A circle 0.85 pixels thick with a diameter of 71 pixels (4500 m) has been generated. It has been turned 90 degrees for presentation so that the zeroes (non-detector) and ones (detector) are on edge. The circle has been artificially squashed by being printed with numeric symbols, and it has been adjusted to fit unequal pixel sizes. This detector tests for valley low points in 124 different directions, and was used to generate the valley lows plotted in Figure 6.*

The NODES program has also been used as an automatic lineament detection system applied to remote sensing images. A large number of standard edge and line detection routines have been added to it including the Sobel, Prewitt, Prewitt variation, Frei-Chen, Kirsch, Line, and Moore and Waltz line detectors. These 3 by 3 filters preferentially detect features in one specific direction. The lineament nodes can be added to each other in order to build up lineaments using a wide variety of detector types and directions. In addition, the new, powerful expanding detector can be directly used to find wider, more major gradients, as well as lighter and darker lineaments in an image. A future application of NODES will be to delineate cultural features such as roads and airport runways on digital images. The size and nature of the expandable detector can be specifically targeted at specific man-made features in order to automatically delineate them for other GIS packages and applications.

## Application to Nisqually valley drainage system, Washington State

Better definition of the seismic activity of western Washington (Figure 2A) is of crucial importance. Recent studies have indicated that this region has a potential for significant earthquakes. Rieken and Thiessen (1990) and Thiessen and Rieken (1991A, 1991B) have identified the Nisqually valley as a probable active fault based on several lines of evidence. A raised relief model of western Washington (1:250 000 scale) was covered with a uniform grey and then illuminated from eight directions. Six or seven interpreters picked



**Figures 2A - 2D:** *Study area of the Puget Sound region, Washington state, for the SEISPLN and LINEAMENT studies. B = Mt. Baker, G = Glacier Peak, R = Mt. Rainier, S = Mt. St. Helens, A = Mt. Adams, O = Olympic Mountains. Figure 2B. High confidence lineaments from analyses of regional elevation models with eight illumination directions and six or seven analysts per image. This figure shows features identified by at least three interpreters on at least three different images that fell within 5 degrees and 5 km of each other. NR is the Nisqually River valley structure. Figure 2C. Seismic foci used for the SEISPLN study, from the Washington Regional Seismic Network (Steve Malone, pers. comm., 1990). Shallow events (less than 35 km) with magnitudes over 0.6 and location qualities of A were used. Figure 2D. Alignment of 32 earthquakes delineated by the SEISPLN program package, and the planar surface fit to the data points which parallels the Nisqually Valley structure.*

lineaments off each image. These were digitized and registered. Compilation maps of lineaments picked by multiple interpreters on each image were prepared using LINEAMENT. Then, high confidence lineaments (picked by multiple interpreters on multiple images) were generated. Figure 2B shows only those lineaments that have been picked by at least three interpreters on at least three different images (within 5 degrees and 5 km of each other), and one of these matches the Nisqually River valley (NR on Figure 2B). SEISPLN was run on the best located (class A) earthquakes shallower than 35 km (Figure 2C) obtained from the Washington Regional Seismic Network (Steve Malone, pers. comm., 1990). An alignment of 32 earthquakes was identified, as shown in Figure 2D. Magnetic anomalies (USGS, 1977) parallel the valley and geologic maps (Schasse, 1987A, 1987B) show discontinuity and apparent offset of lithologic units across the valley. A previous review of seismotectonic features of the Puget Sound (Gower and others, 1985) does not identify this structure. An analysis of the

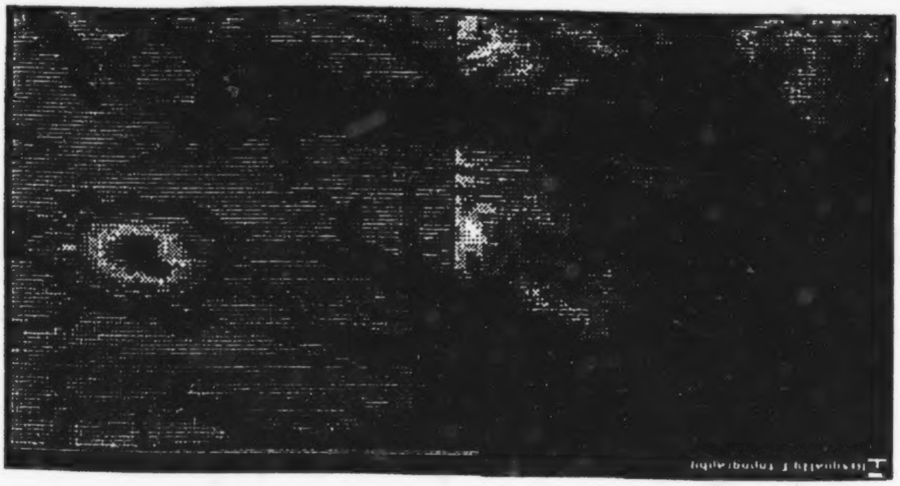


**Figure 3**: *DEM of the eastern half of the Nisqually River area. Area extends from 46 40' N to 47 N latitude and approximately 121 40' W to 122 20' W longitude. Mt. Rainier is the major peak in the right hand portion of the DEM. The Nisqually valley extends to the west-northwest across the lower third of the DEM. Grid spacings are not equal in the east-west and north-south directions (63.16 m versus 92.61 m) and so features are squashed in the NS direction.*

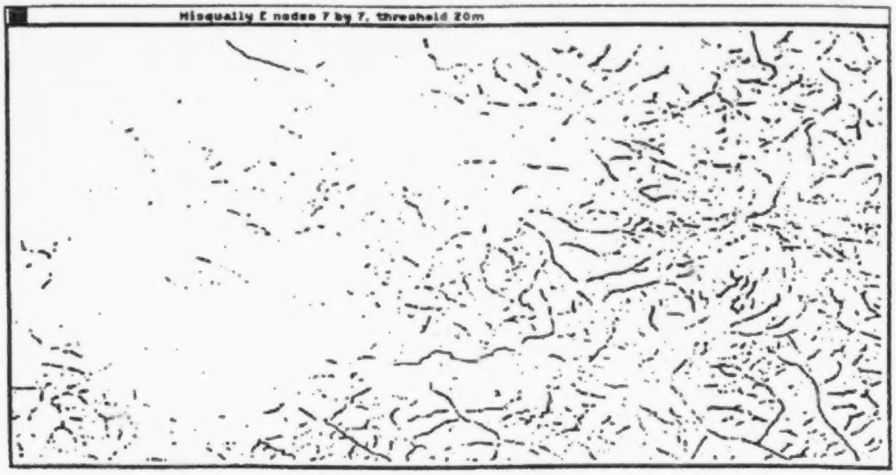


**Figure 4**: *Valley low points detected with the NODES program for the area shown in Figure 3 using a detector 7 pixels (442 m) in size. Valleys had to be at least 20 m deep in order to be displayed. Gray shades indicate the number of directions any grid point was detected (out of a possible of 11 directions).*

topography in order to determine possible fault controlled valleys is therefore warranted.

The DEM used for the study is a mosaic of a portion of two DMA 1:250 000 scale DEMs, the Hoquiam E and Yakima W quadrangles. The study area covers the area of 46 40' N to 47 N latitude and 121 40' W to 123 W longitude. Grid point sizes were 3 arc second by 3 arc seconds, which converts to approximately 63.16 m in the east-west direction and 92.61 m in the north-south direction. The DEM size is 1601 by 401 pixels. The eastern half is shown in Figure 3. Mt. Rainier is the major peak in the right hand portion of the plot. The unequal grid sizes has created an apparent squashing in the north-south direction.

The valley node points that are detected are highly dependent upon the size of the detector being utilized. Figure 4 shows valley nodes detected with a kernal 7 pixels (442 m) in size. Because of the uneven spacing of the DEM grid points (63.16 m versus 92.61 m), the kernal was actually only 5 pixels high (463 m) in order to maintain an approximately circular shape. The circle was 0.85 pixels thick. In addition, the plot of the valley nodes is squashed in the north-south direction, as will all subsequent plots. Valleys had to be at least 20 m deep in order to be designated on Figure 4. Most of the upland drainage systems are quite well designated with good continuity of the individual valley segments. Junctions between valleys are sometimes more complex zones that do not show up on the diagram. The major valleys flowing westward across the central and western portions of the study area are dramatically missing. In fact, these valleys are in zones where there is a total lack of detected valleys. The valley floors are wide enough that, on a scale of 450 m, there is just not enough elevation difference (in excess of 20 m) to the valleys.

The NODES program was run a second time, with a somewhat larger detector, 17 pixels ((1074 m) in size (Figure 5). The detector was again compressed in the north-south direction to 13 pixels (1204 m) in order to approximate a circle. The detector circle was 0.85 pixels thick. Each valley node point could be detected for a maximum of 30 different directions. A color version of Figure 5, with the colors representing the number of times each valley node was detected, reveals some structure to the valleys. The centers of the valleys tend to be detected a larger number of times. The threshold depth used with Figure 5 was 48 m. This means that the valley slopes that are being tested are identical as those used in the preparation of Figure 4. The aspect ratio of the V shaped bottom of the valley is the same for both detectors. A comparison of Figures 4 and 5 shows that there are some different valleys, particularly upland valleys that were found with the narrower detector, but not with the wider. The middle portions of the stream valleys are better formed with the wider detector (Figure 5). The detected valleys are also thicker because the wider detector will see the valley bottom over a wider zone. However, the major trunk streams are still undetected.

Figure 6 shows a wider detector applied to the DEM of Figure 3. The detector itself is shown on its side in Figure 1. The circle is 0.85 pixels thick, has an east-west

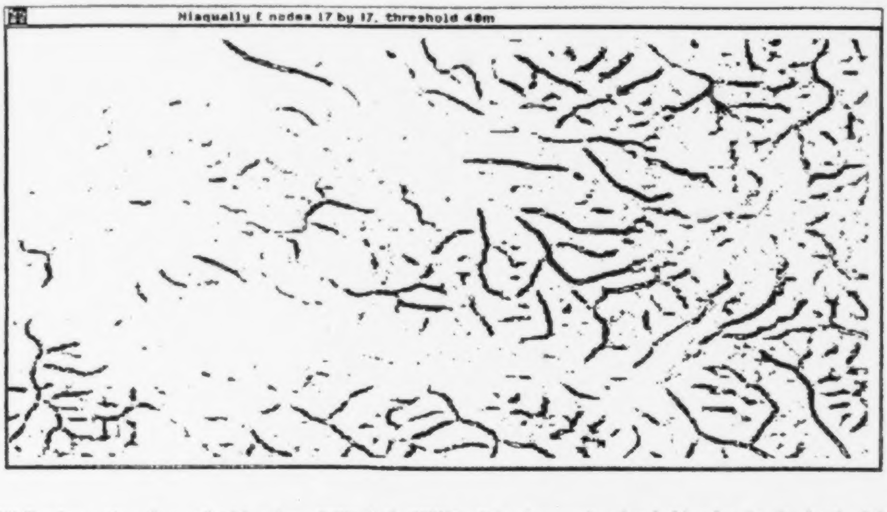


**Figure 5:** *Valley low points detected with a kernel 17 pixels (1074 m) in size, and a threshold value for the depth of the valleys of 48 m. This size and depth yields the same aspect ratio to the valley detector as was used in Figure 4. A maximum of 30 directions were tested when looking for valleys, and gray shades indicate the number of times each node point was flagged as a valley. Note that the centers of valleys are detected a higher number of times. See Figure 4 for more comments.*

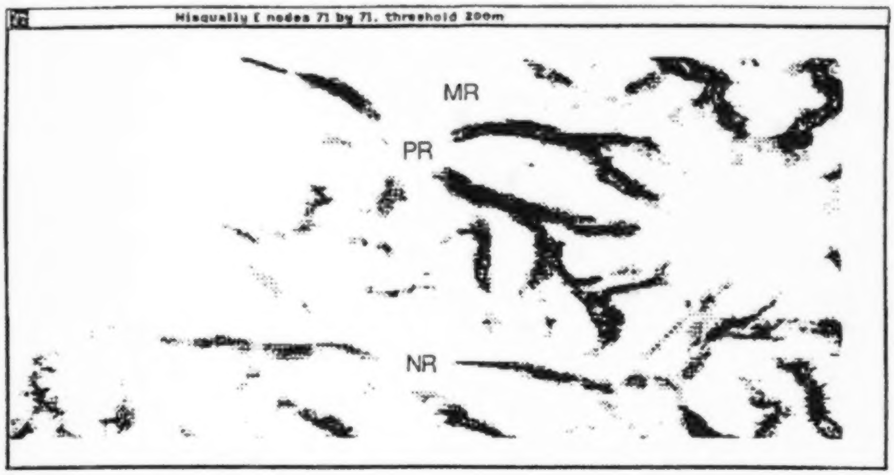


**Figure 6:** *Low points found with the 71 pixel (4484 m) detector shown in Figure 1. Threshold depth was 200 m and the valley aspect ratio is same as for Figures 4 and 5. 124 directions were tested. Major drainage trunks, including the Nisqually River are delineated on this plot, but not on Figures 4 and 5 with smaller detectors.*

diameter of 71 pixels (4484 m) and a north-south diameter of 49 pixels (4538 m). Any valley node can be detected 124 times. Color versions of this diagram illustrate the complex nature of the detections, which can be seen in the gray shade variations of Figure 6. The centers of valleys are found more times by NODES. Straighter valleys (with possible geologic control) tend to be found more times as well. The knowledge of the number of detections at each node point is a powerful new tool for GIS applications. First, it gives operators capabilities to automatically examine the geomorphology of stream drainages. Second, one can density slice the data (show only those nodes that are detected above some critical number of directions, such as 10) and work with only the best formed valleys. This also allows GSA operators to narrow down the wider valleys that are evident in Figure 6 to a more reasonable number of points, or even one pixel wide. When thinning the valleys of Figure 6, one can simply save the node point that was detected the most number of times. For most GSA analyses, these thinned nodes are connected into valley segments and vectorized for input into either the COPLANE or SEISPLN analysis packages.

For Figure 6, the 200 m threshold depth utilized was applied to a distance of 71 pixels. This is approximately equal to the 48 m rise over 17 pixels for Figure 5, and also to the 20 m threshold depth over 7 pixels used for Figure 4. Therefore, the aspect ratio of the valleys being tested is the same for all three examples. If one were dealing with idealized V shaped valleys with a sharp bottom and uniformly sloping sidewalls, then all three figures should display the exact same valleys. In contrast, there is a significant difference in the three displays. The smallest detector found more valleys in upland areas, where the valleys are relatively narrow overall. The intermediate detector (Figure 5) tended to define the middle courses of the valleys. The major drainage trunks of the Nisqually (NR), Puyallup (PR) and Mowich (MR) Rivers were only delineated with the 4500 m detector. The Nisqually River valley is the two elongate These major valleys have relatively flat bottoms and steeper sidewalls, and so cannot be found with the smaller detectors.

There are several possible causes for the anomalous valley shapes. One is that the valleys are simply widening out under natural floodplain development. Maps of the extent of the most recent major mountain glaciation to affect Mt. Rainier and its flanks (Crandell, 1969A, 1969B) show that the glaciers extended through the flat areas in the Puyallup and Mowich River valleys. The glaciers' scouring action would have generated U shaped valley bottoms, with more gentle valley floors and more steep side walls. These U shaped valleys would be difficult to find with narrower detectors, but would be obvious to wider detectors. The Nisqually River glacier extended only as far as the towns of Ashford and National. These are at the western end of the eastern segment of valley nodes delineating the Nisqually valley (Figure 6). Apparently this zone of lows is due to the mountain glaciation. An examination of a geologic map (Schasse, 1987A) shows that the Vashon Icesheet, a continental ice sheet that carved out the Puget Sound, extended towards the Nisqually valley. Its easternmost extent is at the western limit of the flat bottomed

Nisqually River valley segment (Figure 6). Apparently this glacial lobe backed up the Nisqually, causing the valley to be filled with flat sediments. All of these valleys would also be partially filled by stream sediments and glacial outwash. In addition, major volcanic mudflows eminate off Mt. Rainier, such as the Mudflow at Paradise Park (Crandell, 1969A, 1969B) and the Electron Mudflow (Schasse, 1987A)

SEISPLN was run on an integrated data base consisting of the earthquake foci and the end points of the valley segments defined using the 71 pixel detector (Figure 6). This is a true 3D GIS application combining multiple data bases and formats into a spatial analysis. The Nisqually River was a strong planar alignment. First (planar) and second (curved) surfaces were fit to the data. 3D oblique views of the surfaces in combination with the data points were prepared, as were stereopairs. Future research will include the combination of the high quality lineaments (Figure 2B) into the SEISPLN run as well.

## Conclusions

The Geologic Spatial Analysis (GSA) program package has been shown to be a valuable three-dimensional GIS tool. It integrates multiple data bases and formats into a spatial analysis tool that will determine the location of geologic fault and fracture planes in 3D. These are vital for the management of such resources as groundwater, petroleum, and mineral deposits. Fracture systems will affect the environmental impact of waste sites. Active faults, as indicted by the presence of earthquake activity, become crucial in the siting of seismically sensitive buildings. GSA analyses in the past have typically analyzed coplanar alignments of stream valleys, as these are often controlled by geologic structures. A new valley detection routine with an expandable kernal has been formulated. It superimposes a circular detector with a user specified size and thickness onto a DEM. Any point under the circle is tested against the grid point in the center of the circle as well as the symmetrically opposite point across the circle. If both circle points are lower than an input threshold depth, then a valley detection is indicated. For larger detectors, over a hundred different directions may be tested for valleys. The algorithm returns the number of directions the valley node point was detected. The technique was tested on the Nisqually River valley drainage system of western Washington state. This application has shown that the valleys are very scale dependent. Upland valleys are detected with smaller kernals, and major trunk streams. such as the Nisqually River valley, are found with wider detectors. This is true even when one is looking for V shaped valleys with the same slope to the side walls, i.e. the valleys have the same aspect ratio. Minor valleys will exhibit a V shape, but larger valleys will have flatter bottoms due to glacial scouring and/or sediment deposition. These may present problems to some of the standard drainage network programs.

The implications to future GIS analyses of this new valley detector is that major trunk streams can be directly located without having to determine an entire drainage basin topology. In addition, the geomorphology of the valleys, i.e. how well they are formed, is indicated by the number of times the feature was detected. The number of detections can be used to examine only those valley nodes found a large number of times. The size, depth, and minimum number of detections input into the program allows one to design a detector for specific valley shapes. Finally, this detector will become a general tool for any GIS application looking at streams and or valleys.

## Bibliography

Band, L.E., 1986, Topographic Partition of Watersheds with Digital Elevation Models, Water Resources Research, v. 22, p.15-24.

Beaver, D.E., Eliason, J.R., and Thiessen, R.L, 1989, Regional Digital Analysis of Major Crustal Structures in Washington State, Proceedings of the Seventh International Conference on Basement Tectonics, in press.

Crandell, D.R., 1969A, Surficial Geology of Mount Rainier National Park, U.S. Geological Survey Bulletin 1288, 41 p.

Crandell, D.R., 1969B, The Geologic Story of Mount Rainier, U.S. Geological Survey Bulletin 1292, 43 p

Desachy, J. and Lammali, N., 1989, Automatic Hydrologic Network Computation from Digital Elevation Model, 1989 International Geoscience and Remote Sensing Symposium (IGARSS'89), p. 1596-1599.

Eliason, J.R., 1984, A Technique for Structural Geologic Analysis of Topography, Ph.D. thesis, Washington State University, Pullman, WA, 166 p.

Eliason, J.R. and Eliason, V.E., 1985, A Comparative Study of Fracture Planes Computed from Topography and Lineaments from Imagery with Structures and Mineralization in the Magnesite Belt of Washington State, International Symposium on Remote Sensing of the Environment, Fourth Thematic Conference, Remote Sensing for Exploration Geology, p. 655-664.

Eliason, J.R. and Foote, H.P., 1986, A Digital Topographic Analysis Technique for Detection of Crustal Fractures, Proceedings of the Sixth International Conference on Basement Tectonics, p. 198.

Eliason, J.R. and Thiessen, R.L., 1987, Geologic Spatial Analysis, a New Multiple Data Source Exploration Tool, International Symposium on Remote Sensing of the Environment, Fifth Thematic Conference, Remote Sensing for Exploration Geology, p. 763-774.

Foley, M.G., Beaver, D.E., Glennon, M.A., Mroz, T.H., Eliason, J.R., and Thiessen, R.L., 1988, Applition of Remote Geologic Analysis to Gas Exploration in West Virginia, Geological Society of America Abstracts with Programs, v. 20, p. A229.

Gower, H. D., Yount, J. C., and Crosson, R. S., 1985, Seismotectonic Map of the Puget Sound Region, Washington, 1,250,000, U.S. Geological Survey Miscellaneous Investigation Series, Map I-1613.

Greenlee, D.D., 1987, Raster and Vector Processing for Scanned Linework, Photogrammetric Engineering and Remote Sensing, v. 53, p. 1383-1387.

Hadipriono, F.C., Lyon, J.G., Li, T., and Argialas, D.P., 1990, Photogrammetric Engineering and Remote Sensing, v. 56, p. 905-909.

Haralick, R.M., 1983, Ridge and Valley on Digital Images, Computer Vision, Graphics and Image Processing, v. 22, p. 28-38.

Jenson, S.K., 1985, Automated Derivation of Hydrologic Basin Characteristics from Digital Elevation Data, Proceedings of Auto-Carto 7, Digital Representations of Spatial Knowledge, p. 301-310.

Jenson, S.K. and Domingue, J.O., 1988, Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis, Photogrammetric Engineering and Remote Sensing, v. 54, p. 1593-1600.

Johnson, L.K., 1988, Computer Analysis of Remote Sensing and Geologic Datasets, M.S. thesis, Washington State University, Pullman, WA, 215 p.

Lammers, R.B. and Band, L.E., 1990, Automating Object Representation of Drainage Basins, Computers and Geosciences v. 16, p. 787-810.

Morris, D.G. and Heerdegen, R.G., 1988, Automatically Derived Catchment Boundaries and Channel Networks and Their Hydrologic Applications, Geomorphology, v. 1, p. 131-141.

O'Callaghan, J.F. and Mark, D.M., 1984, The Extraction of Drainage Networks from Digital Elevation Data, Computer Vision, Graphics, and Image Processing, v. 28, p. 323-344.

Palacios-Velez, O.L. and Cuevas-Renaud, B., 1986, Automated River Course, Ridge and Basin Delineation from Digital Elevation Data,

Journal of Hydrology, v. 86, p. 299-314.

Pitz, C. and Thiessen, R.L, 1986, Lineament Analysis and Structural Mapping of the Trans-Idaho Discontinuity and their Implications for Regional Tectonic Models, Proceedings of the Sixth International Conference on Basement Tectonics, p. 16-24.

Peucker, T.K. and Douglas, D.H., 1975, Detection of Surface-Specific Points by Local Parallel Processing of Discrete Terrain, Computer Graphics and Image Processing, v. 4, p. 375-387.

Qian, J., Ehrich, R.W., and Campbell, J.B., 1990, DNESYS - An Expert System for Automatic Extraction of Drainage Networks from Digital Elevation Data, IEEE Transactions on Geoscience and Remote Sensing, v. 28, p. 29-45.

Rieken, E.R., 1985, Computer Generated Fault Surface Determinations from Earthquake Foci, Washington State, 1969-1983, M.S. thesis, Washington State University, Pullman, WA, 126 p.

Rieken, E.R., Rohay, A.C., and Thiessen, R.L., 1991, Analysis of the Cascadia Subduction Zone Using Alignments of Seismic Foci, Western Washington, Geological Society of America Abstracts with Programs, v. 23, p. 92.

Rieken, E. R. and R. L. Thiessen, 1990, SEISPLN Analysis of the Juan de Fuca Plate Beneath Western Washington, v. 15, Geologic Association of Canada - Mineralogical Association of Canada, Program with Abstracts annual meeting, p. A111.

Rieken, E. and Thiessen, R., 1991, The Geometry of the Portion of the Juan de Fuca Plate Beneath Western Washington as Determined by Hypocenter Alignments: EOS, v. 72, in press.

Rieken, E.R. and Thiessen, R.L., in prep, Three-Dimensional Model of the Cascadia Subduction Zone Using Earthquake Foci, Western Washington, Bulletin of the Seismological Society of America, in prep.

Schasse, H.W., 1987A, Geologic Map of the Centralia Quadrangle, Washington, Washington Division of Geologic and Earth Research Open File Report 87-11, 27p., 1,100,000 scale.

Schasse, H.W., 1987B, Geologic Map of the Mount Rainier Quadrangle, Washington, Washington Division of Geologic and Earth Research Open File Report 87-16, 43p., 1,100,000 scale.

Smith, T.R., Zhan, C., and Gao, P., 1990, A Knowledge-Based Two Step Procedure for Extracting Channel Networks from Noisy DEM Data, Computers and Geosciences, v. 16, p. 777-786.

Thiessen, R.L., 1986, LINMNT, Lineament Analysis, National Association of Geology Teachers Special Publication, Computer Software Designed for use in Undergraduate Education, p. 183-187.

Thiessen, R., 1991, Evaluation of Clustering Methods Applied to Fracture Trends: EOS, v. 72, in press.

Thiessen, R.L., Eliason, J.R., Beaver, D.E., and Watkinson, A.J., 1989A, Determination of Crustal Fractures in Northeastern Washington Using Geological Spatial Analysis and Their Correlation to Mineralization, Geological Society of America, Abstracts with Programs, v. 21, p. 151.

Thiessen, R.L., Eliason, J.R., Brougher, C.W., Johnson, K.R., Mohl, G.B., Foley, M., and Beaver, D.E., 1988, Structural Analysis of the Central Columbia Plateau Utilizing Radar, Landsat, Digital Topography, and Magnetic Data Bases, Eighth International Conference on Basement Tectonics, Program with Abstracts, p. 33.

Thiessen, R.L., Eliason, J.R., and Johnson, K.R., 1989B, A Review of Techniques for Identification of Faults on Acoustic Images for use in Geologic Spatial Analysis, International Symposium on Remote Sensing of the Environment, Seventh Thematic Conference, Remote Sensing for Exploration Geology, 955-968.

Thiessen, R.L., Eliason, J.R., and Rieken, E.R., 1989C, Three-Dimensional Computer Analysis and Modelling of Remote Sensing-Structural Geologic Problems, 1989 International Geoscience and Remote Sensing Symposium (IGARSS '89), p. 89-92.

Thiessen, R.L., Eliason, J.R., and Rieken, E.R., 1989D, Three-Dimensional Computer Analysis Applied to Exploration Problems, Invited presentation at International Symposium on Remote Sensing of the Environment, Seventh Thematic Conference, Remote Sensing for Exploration Geology.

Thiessen, R.L., Johnson, L.K., Foote, H.P., and Eliason, J.R., 1987, Surface Reflectance Correction and Stereo Enhancement of Landsat Thematic Mapper Imagery for Structural Geologic Exploration, International Symposium on Remote Sensing of the Environment, Fifth Thematic Conference, Remote Sensing for Exploration Geology, p. 763-774.

Thiessen, R.L., Kimball, C.S., Rieken, E., Weberg, E., and Barsotti, A., 1985, Comparison of Lineament Studies Using one Thematic Mapper Image and Using a Multi-Platform Approach, Sixth International Conference on Basement Tectonics Abstracts with Programs, v. 6, p. 37-38.

Thiessen, R.L. and Rieken, E.R., 1991A, Integration of Seismic Foci, Digital Topography, and Lineaments to Define Three-Dimensional Crustal Fractures in Western Washington, Geological Society of America Abstracts with Programs, v. 23, p. 103.

Thiessen, R.L. and Rieken, E., 1991B, Delineation of Three-Dimensional Crustal Fault Systems in Washington Using Computer Automated Techniques: EOS, v. 72, in press.

Thiessen, R.L., Watkinson, A.J., Eliason, J.R., and Ader, M.J., 1989E, Review of the Relationship of Fractures to Stress Systems and Their Implications for Remote Sensing and Geologic Spatial Analysis, Summaries, International Symposium on Remote Sensing of the Environment, Seventh Thematic Conference, Remote Sensing for Exploration Geology, 1331-1332.

Tribe, A., 1991, Automated Recognition of Valley Heads from Digital Elevation Models, Earth Surface Processes and Landforms, v. 16, p. 33-49.

USGS, 1977, Aeromagnetic Map of Northern and Eastern Parts of the Puget Sound Area, Washington, U.S.G.S. Open File Report 77-34, 1,250,000 scale.

Van Pelt, J., Woldenberg, M.J., and Verwer, R.W.H., 1989, Two Generalized Topological Models of Stream Network Growth, Journal of Geology, v. 97, p. 281-299.

GIS
92

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# A decision support system using neural nets and genetic algorithms for multi-objective State Park management

**Randy Gimblett**

Department of Landscape Architecture
College of Architecture and Planning
Ball State University Muncie, IN 47305
Tel: 317-285-8065

**George L. Ball**

Advanced Resource Technology laboratory
School of Renewable Natural Resources
University of Arizona Tuscon, AZ 85721

## Abstract

*Traditional geographic information systems (GIS) have provided a mechanism to specifically decide in a linear fashion the individual weightings as to the relative importance of physical variables for single spatial modeling applications. However, as the sophistication in the decision-making process increases the examination of many more complex resource and non-resource relationships becomes virtually impossible.*

*What is required is the inclusion of intelligence into the sophisticated modeling structure that allows for multiple decision-making and concurrent adaption to evolving conditions. In addition, methods must be developed to effectively include the public as a resource and subsequently public values into the park management process.*

*The purpose of this paper is to describe a prototype decision-support system for multi-objective state park management using neural network and genetic algorithm technology. This paper will examine and contrast how the resources of Mounds State Park in Indiana are currently being managed, how the park according to State policy is supposed to be managed and more importantly how the public perceives the resource management of the park. By understanding how the experts manage the park and contrasting how the public perceive such management objectives, the net and GA based model provides the resource managers with a tool to examine conflicts and tradeoffs over time for long term planning and management strategies which incorporate public values. Differences between traditional GIS and neural network/GA approaches will be presented.*

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Parser design strategies for GIS natural language interfaces

**G. William Lydiatt**

President, Interface S.A.
Santa Lucia, OF 71
Santiago, Chile

## Abstract

*The only useful database is an accessible one, and accessibility can be measured as a function of the user's ability to express requests for raw or processed information with satisfactory results.*

*In as much as any given database has a fixed or proprietary structure, it is possible to apply well known natural language parsing techniques that can make a deep analysis of queries, and elicit the desired response or activity.*

*A number of GIS systems, however, work over off-the-shelf or parametric databases, whose structures are application dependent, and which vary according to what is currently loaded.*

*The situation is made more complex by the requirement that the parser be able to execute functions, not only over the database itself, but also within the GIS host environment. In such cases, the number of rules required to adequately communicate with both the GIS and the database can greatly outweigh the benefits of applying natural language techniques to the case.*

## Introduction

This paper reviews some of the major issues in natural language parser design based our experiences in developing a natural query language over Dbase III compatible databases, for Spanish language users. The resulting strategy is based on the concept of "island driving" or partial parse paths that generate rules and clauses that can be subsequently executed by a clause processor written. Both the parser and processor are written in Prolog.

## Parser design strategies for GIS natural language interfaces

All natural language parser design quite properly begins with a simple question: what is it going to be used for? The answer seems obvious: to communicate with GIS systems and their associated databases. The next question is how.

## Full or partial parse

The depth of linguistic analysis that will be required is an issue that must be clear from the beginning. The literature of computational linguistics and natural language processing (NLP)is replete with sophisticated parsers, each of which has been designed to evaluate language from a certain standpoint or for a certain purpose. As a result, techniques that have been used, for example, to write summaries of newspaper articles, only too frequently have little to offer for application specific database queries, which, one day might involve elevation and core habitats, and the next day, schools of fish.

In general, query flexibility must pay a high price in profundity of analysis. On the other hand our, experience teaches that it is more important that the parser has a natural feel than that it makes a complete syntactic analysis.

## Practical platform constraints

All computer software is built for one or more specific platforms, and is defined by the platform itself, the task at hand, marketing considerations, cost, and others. In the case of a modest project such as NaturalQl, we assumed limited available end-user resources (Intel 80C88 - 286 512-640 KB RAM and disk. This self-imposed restraint was based on the reality of third world markets.

Outside North America and Western Europe the majority of countries that can, and should take advantage of both GIS and NLP technology, are only recently entering the 386, Windows, Unix networked world. The acquisition of computer systems of any type still remains a major long term investment in the third world, and buyers are relatively slow to be affected by current hardware fashions. A serious GIS system can easily cost the equivalent of more than ten man years of manual work. Since the platform is the market, the cheaper the better.

On the other hand, it is difficult to assume less that a 386 platform, with memory caches, co-processor and the like, given both the normal GIS processing requirements plus the natural language overhead.

## Built in or on top

Another critical design issue is whether to integrate the parsing function within an existing GIS product, or add it on top. The former can require redesigning the whole product but produce spectacular results, whereas the latter can mean sacrificing a certain degree of functionality for relative ease of implementation. It is a classic cost-benefit problem, which, as discussed below, is somewhat aggravated by the nature of modern databases.

## Language of implementation

Artificial intelligence professionals are divided into three camps, object oriented C under Windows, North Americans LISP workstations, and Japanese and European Prolog PC's and supercomputers.

In our case, Prolog was the language of choice. Prolog directly implements classic first order predicate logic, which is based on the basic human logical axioms common to both East and West for thousands of years. The language incorporates backtracking mechanisms in such a way that the task is more one of logically expressing a process and providing adequate control than the design and implementing custom made search and unification algorithms.

Prolog is an ideal medium for evaluating the validity or invalidity of sentences and arguments and, for a symbol-based system that may require hundreds of rules, a stable, well known underlying logic is essential.

In addition, Prolog's parallel processing capacity, which, although it can lead to some interesting backtracking bugs and control problems, is ideally suited for maintaining the large number of pointers typically required by symbol processing.

On the down side, Prolog is a weak cousin to C for implementing other interfaces, such as with digitizers, graphics printers, etc. As such, it is not without merit to consider using C for the spatial graphics, driven by Prolog or Lisp parsers and inference engines.

## Domain issues

Any GIS natural language interface must by definition be capable of handling both attribute and spatial data, separately and together.

This is perhaps the key design issue for GIS systems, in as much as GIS parsers must be both domain specific and application independent. This seeming paradox is resolved by conceiving of GIS systems as deterministic qua geography but non-deterministic qua information systems. The domain is clearly defined by its I/O technology, and incorporates a set of predetermined conceptssuch as layering and data types. On the other hand it is non-deterministic with respect to individual applications, which can include those, that at first glance, can appear to have little to do with geography as we commonly understand it.

Even more important for any discussion of natural language, is the fact that GIS is data-driven, and that outside tne data type set, the domain does not impose a predetermined structure on the representation of data. For example, the number of entities that must be managed by a natural language system can vary widely from application to application, and entity labels are effectively parametric. As shown below, this has an important effect on parser design.

To a very considerable degree, the split personality of GIS is directly associated with its spatial and attribute functionality, in as much as application specific data structures are enveloped by certain predetermined *'a priori'* conditions.

## *A priori* constraints and issues

A priori constraints on data are provided by both GIS hardware and software, the former through forms of graphic representation, data capture and output, and the latter by tailoring and organizing the data for delivery by the hardware. Natural language systems can take advantage of this by "hard-wiring" the syntactic elements associated with commands and interrogatives directly to these pre-established functions and concentrating instead on data access, retrieval and delivery.

Unfortunately, the 'traditional' parsing strategies such as augmented transition networks' (ATN), bottom up, charts, etc. rest on a basic assumption, that is, that there is a clearly defined, finite and accepted syntax in language. Although this is most definitely the case in human languages, else communication would be impossible, each application specific data representation presents new syntactic challenges upon loading it into the machine.

The "traditional" parsing strategies encounter serious impediments when faced with completely parametric data base structures that can change from moment to moment according to the application.

Clearly, the basic vocabulary (entity labels), ordering, conceptual grouping and content (semantics) radically differ from domain to domain, for example between forest and land management, noise pollution and submarine resources.

What is given, however, is a common set of data types, and from this base other parsing strategies can emerge.

## Island parsing

To appreciate how non-determined database structures rapidly escape formal fixed syntax parsers, consider the following example:

Although disjunctives and conjunctives are basic to all computing processes, they present considerable problems when trying to provide for maximum flexibility when querying databases of undetermined structure, for example to provide for the possibility of user-defined ordering of the input sentence.

First consider the following command:

show land cover

This can be easily expressed in rules such that:

query = command, object

and psuedocodified as follows:

```
display:-
land_cover.
land_cover:- land_cover_objects(X).
land_cover_objects(1).
land_cover_objects(2).
.
.
.
land_cover_objects(n).
```

The command, of course, can be "wired" to spatial data representation processes, based on fixed conditions and activities. For example, words such as overlay, plot, display or their user-defined synonyms could be initiate specific activities valid within the GIS domain.

Note, however, that this grammar admits only two valid terms:

1 show, which is a command that will initiate a specific system function, and
2 land cover, a set of objects of the class land_cover.

To extend the example consider the following query:

show "land cover and utilization within 500 meters of core habitats"

which must by implication provide parsing rules for the following:

show land_cover

or

query = command,object

show land_cover, utilization

query = command,object,object

show land_cover, utilization, core_habitats

query = command,object,object object

show land_cover, <, 500_ meters

query = command,object,operator_expression,object

show utilization, <, 500_ meters

query = command,object,operator_expression,object

show land_cover,utilization,<,core_habitats

query =
command,object,object,operator_expression,object

where in order to be truly flexible, an object can be any one of:

land_cover,

utilization,

core_habitats

and each object can be a class of other objects.

Finally, consider:

show "land cover and utilization within 500 meters of corehabitats" and elevation above 450 meters"

which adds several other valid sentence types to the ones described above, including:

query =
command,object,object,operator_expression,object,
operator_expression object

It is easy to see maintaining flexibility in analysis with a traditional parser can lead to a requirement for hundreds if not thousands of rules that consume system resources, and require extensive periods of time to execute.

This is precisely the problem we faced in our attempt to build a useful, practical interface for Dbase, with top down left to right parsing. The end result that was every time we changed the database we had a new situation with a new basic vocabulary and new demands for syntax analysis.

Very early in the project it became evident that it was necessary to abandon any idea of a formal parser and sacrifice completeness and depth for practical access.

Within the context of GIS, one response to this situation is to equip the parser with a limited set of priori domain rules that are compatible with command functions of the GIS hardware and software, such that, for example, all commands passed to the parser will execute system functions.

The remaining content of a query can then be evaluated as clauses of one or more words, with or without incorporated conditions. This dramatically reduces the number of rules required, in effect to only those that are necessary to manage all data types, all conditions and all desirable extensions of conditions (for example handling dates, or concepts such as "approximately' or "about"). In addition such a parser must be able to associate words and/or phrases with data items and/or sets.

In practice, the resulting inference engine is more a clause processor than a parser, and, as such, can take advantage of capability of Prolog to generate its own rules as required during execution, a technique implemented by means of meta-programming.

## Islands

All but the most rigorous parsers filter a significant portion of the input query, including punctuation, non functional words, and others.

In addition, such a parser can provide for very flexible user-based assignation of synonyms and associations including:

- single words to single data items
- single words to multiple data items
- phrases to single data items
- phrase to multiple data items

This permits grouping and expressing whole sets of data as one word or phrase, such as "land cover" or "core habitats, and allows users to make commands like "overlay land cover", thereby initiating a complex process within the domain in a completely natural way.

Such associations and groups are asserted to an internal database as clauses, that are subsequently fired as encountered or required.

Mathematical operators and other conditions are also asserted as clauses, however in a somewhat different format. These clauses necessarily include information about both the subject and object of the operation as well as the operator itself, and can also provide for more ambiguous concepts such as approximately, near,etc.

This parsing strategy has been called "island driving" in that the parser works from islands, such that it can find a partial parse path at any point over the data and resolve the "islands" expressed in the query. More formally, it can be said to be a data driven, partial parse clause processor with sub-nets.

User defined commands and interrogatives are "hard wired" to domains commands, and ,as required, islands discovered in the input text are provided with the capacity of being evaluated to both the right and left. The resulting clause is then asserted to the internal database where it waits for processing by the clause processor.

As applied to GIS, such a system might assume that "command" phrases have highest precedence, in that they are required to determine logical output devices, instruct the parser where to look for data, etc. Obviously a command directed to raster output must be distinguished from one that conditions attribute data.

In those cases where the data required to satisfy the query are mixed or derived from different databases (for example attribute, coordinates) the range of a island might be defined as coveringall terms up to the next command, resulting in commands of the following type:

> get land utilization attribute data, and show coordinates data.

Island parsing routines must provide for analysis in both directions and also process data in fields encountered en route to the edge of the island.

Suppose that such an island is defined as having a range over all objects between a command and a "keyword", and a certain range of objects to the right of the keyword.

For example

|——————left range———|- keyword—|right range|

show merchantable coniferous stands classified by origin!

might

1. assert the command clause at the top of the database, permitting the clause processor to direct output as an overlay,

1. Eliminate the "!" and "by":

|———left range——————|- keyword—|-r. range-|

{display} merchantable coniferous stands classified origin

2. Convert the synonyms

|- l. range - |—keyword—|- r. range—|

{display} MERCHANT CONIF SORT ORIGIN

3. Evaluate the left range as the subject of the operation and the right range as the criteria by which to perform the operation, identify the process required by "keyword or condition, and generate the corresponding Prolog clauses.

4. In the case of field searches, sufficient information must also be asserted in the clause to initiate a search of the appropriate field or fields, instead of simply returning the value associated with a known label.

5. The clause processor then retracts each asserted clause in a preestablished order, evaluates it and initiates the search for the appropriate data.

The benefit of adopting this strategy is that it is capable ofhandling 'n' data items without the requirement of anticipating their analysis by 'n' rules plus overhead. The parser simply executes *a priori* functions and generates the clauses on the fly. These are then called as rules by the clause processor for execution over the database.

On the other hand the island strategy does have a potential down side, It is very stack and heap intensive, and works best with liberal amounts of available memory. In addition it requires very careful handling of control cuts to direct Prolog's backtracking and unification mechanisms and prevent the program logic from blasting off on unexpected search branches.

This island driving concept, augmented by the metaprogramming techniques permitted by Prolog can be considered as a viable alternative for GIS natural language interfaces, and although considerable work can be anticipated in delivering extracted data to the GIS platform, it promises a truly natural feel and functionality.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Phase II EPPL shell: A GIS application for the field resource manager

**Dave Martodam**

Computer User Support Coordinator
Minnesota Department of Natural Resources
Division of Forestry
500 Lafayette Road
St. Paul, MN
USA 55155-4044

## Abstract

*The Phase II EPPL Shell is a PC-based GIS tool that the field natural resource manager can use to access and spatially analyze forest inventory data.*

*The shell is a menu driven interface To EPPL7, a raster-based GIS. The interface allows the user to access, query, and perform specified GIS analysis on forest stand data and any associated forest management prescription data. Color maps of the analysis can be viewed on the screen or printed. The user doesn't need to know how to run the EPPL7 GIS. Instead, the emphasis is on knowing the data's potential uses and limitations.*

*The Minnesota Department of Natural Resources - Division of Forestry is implementing the shell in field offices. It is being used to supplement traditional forest stand level decision making, to plan and do field work more efficiently, and to overcome limitations of the present forest inventory design. The tool is enabling field managers to better resolve forest land management conflicts caused by increasing demands on the forest resource.*

*The shell can be run on a 286 class PC with EGA color monitor and 40 MB harddrive. The recommended PC is a 386 with VGA color and a 80+ MB harddrive. Map output is done on an HP Paintjet color printer. The shell demonstrates the feasibility of developing and using a GIS interface in a field environment. It serves as a stand-alone analytic tool and also has the potential to be incorporated into a larger application or information system.*

## Introduction

*The Minnesota Department of Natural Resources* The mission of the Minnesota Department of Natural Resources (DNR) is "To preserve, protect and enhance Minnesota's natural resource heritage in order to benefit the environment, economy, and quality of life for all Minnesotans, present and future" (Minnesota Department of Natural Resources, 1991). This paper will focus on how the DNR Division of Forestry (DOF) is using GIS technology at the field level to achieve that goal.

There are 14.7 million acres (5.95 million hectares) of commercial (timber producing) forest land in Minnesota. The DOF administers approximately 2.8 million acres of that land. It also provides forest management assistance on approximately 6.2 million acres of private, non-industrial forest land. (Forest Statistics For Minnesota, 1990).

It is the lead forest management agency in the state and develops policy and procedures that guide the management of all the state's forest land.

The forest products industry is the second largest manufacturing industry in Minnesota. It employees 55,600 people and generates an annual manufactured product value of 4.98 billion dollars. (Minnesota Forest Industries Factbook, 1988). The industry is composed of 8 pulp and paper companies, 10 oriented strand board, fiberboard, and related companies, and about 700 sawmills and related wood product manufacturers. The pulp and paper and oriented strand board industries have made approximately 2 billion dollars of capitol investment between 1979 and the present. Planned expansion between now and 1997 is another 2 billion dollars. (Minnesota Forest Resources, 1991). These planned expansions would increase the harvest of Minnesota timber to nearly the maximum sustainable level for some major species.

Environmentalists, preservationists and people within the forestry community alike are concerned about the ability of the state's forest land to support this increased level of use without adverse environmental effects. This concern has manifested itself in events as diverse as the spiking of a stand of pine scheduled for harvest (a first for Minnesota) to a project to produce a "Generic Environmental Impact Statement for Timber Harvest." Minnesota is no different from other states or provinces as it grapples with questions about bio-diversity, old growth, integrated resource management, visual management, and similar issues generated by the ncreased competition for the resource. The situation is further complicated by constantly shrinking budgets which restricts the DNR's ability to meet the increasing demands.

## The phase II forest inventory

The primary database used by DOF field managers to make forest management decisions is the Phase II Forest Inventory. The Phase II inventory is a stand-level inventory. Forest stand boundaries are drawn on current, low-altitude aerial photos. Stand boundaries are delineated within sections of the Public Land Survey. The township is the standard unit that inventory work is done on and maps are developed by township. Field visits are done to correct stand boundaries and record up to 92 attributes for each stand.

The original database design was a computer flat file and mapping system. It is being converted to a GIS database using ARC/INFO (tm) GIS software. The digitzation process is about 75% done for the land the DNR administers. The digital database has four layers; forest cover type, roads and trails, water, and inclusions (small, pseudo-stands within a stand).

The inventory process generates the following products which the field manager uses to make management decisions.

1. Township maps (4 inch to the mile scale) of cover types. The maps may be hand drawn or produced by a plotter if the data is digital.
2. Stand data listings and summary reports.
3. Floppy disks with files containing the stand attribute data.

The primary users of the inventory data are field foresters and wildlife managers. They use it mainly to make stand-level management decisions such as whether a stand should be harvested or reserved. Higher levels of the organization use the data to create summary information used to develop overall management strategies.

New demands on the forest resource have provided opportunities to use the inventory for uses far beyond its original purpose. The designers of the Phase II inventory had no idea that it would be used to address old growth, bio-diversity, visual management and similar issues that face today's resource manager. It is a tribute to the original inventory design that the DNR can use it to deal with those issues as well as it does.

## Phase II EPPL shell development

The field managers have a wealth of information from the Phase II inventory at their disposal. They have paper maps, paper listings of the stand attribute data, and a computer file of the stand attribute data. However, it can be difficult to use the tabular and spatial data concurrently. Managers must look at a stand's attribute data and then at the associated cover type map to use the two data sources at the same time. For example, the attribute data might show a pine is stand is ready for harvest. Then the manager finds the associated map, inspects it, and discovers the pine stand is next to a popular campground and decides not harvest it.

This manual process is laborious at best. It is impossible if many stands are being considered. The computer database can be queried for more complex questions. However, the cover type maps only show the spatial relationships of the cover types, average tree size and density. The maps are not useful if the manager needs to consider forest age or site quality relationships.

The DNR has ARC/INFO and the Phase II database available on Data General UNIX Workstations at its central office in St. Paul. Four regional offices are also installing workstations. Even when these workstations are fully operational, they will provide limited access to GIS for the field manager because of distances from field sites. They certainly can't provide daily support for field management decisions.

The DOF has provided with field offices with microcomputers. While it is technically feasible to install a full-featured GIS in a field office, there are several reasons why it is not possible to do so. The cost of a full-featured GIS software package is still high enough to make it prohibitive for every field station to have one. Learning to use a full-featured GIS would be very expensive also. It is a task that most field managers, who generally have little computer experience and often little aptitude for computers, are not willing to undertake. Also, due to limited resources and the technical complexity of the software, the DOF's information systems staff is not capable of supporting many users on a full-featured GIS.

Many field managers were frustrated that they could not use GIS to get the most out of their Phase II data. They knew that the technology would help them work more efficiently in the face of budget reductions and the increasing demands being placed on the forest resource. To meet this need the DOF's Littlefork field office proposed a pilot project to develop and evaluate a PC-based GIS that would be useable at the field level.

In September, 1989, the DOF management approved and assigned resources for the project. The goal of the pilot project was to test the technical feasibility and utility of a PC-based GIS at a field office. The system would have the following features:

1. Enable access and analysis of Phase II data and any associated stand prescription data.
2. Users of the system would be field resource managers (mainly foresters and wildlife managers).
3. Produce color maps that could be viewed on the screen and/or printed.
4. Have a menu driven user interface and be easy to use.
5. Run in real time on an AT class PC which most field offices already had.

A project team used a structured system development process to manage the project. A written project proposal defined goals, resource needs, and schedules. The project team consisted of a senior field manager who provided leadership and management authority and credibility; a middle management staff forester who did project administration; a contractor who provided technical GIS expertise and directed programming work; and a field forester who represented potential field users. The project also had a corporate sponsor, in this case a Regional Forest Supervisor, who assigned resources and championed the project with top management. The project team worked very closely with the potential users throughout the whole project to determine needs and make sure that the product met them.

The Environmental Planning and Programming Language Version 7 (EPPL7) was chosen as the GIS software to develop the shell in. EPPL7 is a low-cost, full-featured, raster-based GIS that runs on IBM PC's, PS/2's and compatibles using DOS 2.0 and above. It is produced, marketed, and supported by the Minnesota Department of Administration, Land Management Information Center (LMIC). (EPPL7 Version 7 Release 2.0 - USERS GUIDE)

EPPL7 contains a full set of geoprocessing routines partitioned into four modules: EPPL for analysis and file maintenance routines; DISPLAY for screen graphics; DOTPLOT for hardcopy output; and DIGITIZE for vector data entry and modification. EPPL7 successfully integrates vector and raster technologies for the simultaneous display of raster data complete with up to ten vector and/or tag (character) overlays. There are also attribute creation and manipulation routines with interfaces for INFO, ERDAS, GRASS, DXF, and DLG data structures as well as other standard output formats. EPPL7 is also easy to customize using conventional programming languages and 'REPLAY' files (macros that record and playback EPPL sessions).

The DOF contracted with LMIC to do the necessary programming and provide EPPL7 and GIS technical expertise. The Phase II EPPL Shell is written in Borland Turbo PASCAL version 5.1. It has a flexible design that is easily enhanced and modified with little change to the overall system design or flow of control.

The project followed a two stage process. The first stage created a menu driven interface to the Phase II database that allowed the user to access, do interactive "AND" operator queries, and produce maps of the query. It was basically an automated mapper. The users in the pilot area then worked with the interface to learn what the potential use was and determine the specific GIS features they wanted. A process to convert the Phase II ARC/INFO vector data to a raster format usable by EPPL7 was also developed.

The second stage incorporated the GIS features users had decided were the most important to them. It also included a long period of testing to make sure the product was as bug free as possible before it was declared ready to use. Also, a user manual, project documentation, and an implementation plan were developed during this stage.

The pilot project was finished in September, 1991. The two year development time was a result of the two stage, iterative process used to develop the shell, and an extensive beta test period. Also, there was a six month period when no work was done on the project while we searched for funds to continue the second stage. The project cost approximately $20,000.00. This does not include DOF staff time and expenses. Five thousand dollars of the $20,000.00 were used to convert ARC/INFO data to EPPL7 format in the pilot project area.

The final product is the Phase II EPPL Shell program on a floppy disk, a users manual, trained support people, procedures for installing the software and data and training users. It also includes a process for converting ARC/INFO data to the EPPL format needed by the shell.

The minimal hardware configuration the shell will run on is a 286 PC, with 640 K RAM, an EGA color monitor, and a 40 MB harddrive. The recommended configuration is a 386 with math coprocessor, VGA color monitor, and an 80+ MB harddrive. Color maps are printed on a Hewlett Packard Paintjet or Paintjet XL color inkjet printer.

The Phase II EPPL Shell features and performance considerably exceed project expectations. This was due in part to the hard work and innovation of the LMIC consultants who wrote the shell. An example is suggesting the inclusion of 1:100 000 DGL road and water vector overlays which weren't part of the original project specifications. The primary features of the shell available from the main menu are:

1. DIRECTORY - Allows the user to pick a township to work with from a list of townships loaded on the harddisk.
2. QUERY - The most powerful feature. Simple "AND" operator queries can be made on any combination of attribute data. Complex queries are created using Boolean logic. They can be edited and stored in query logic files. Query logic files can be developed into complex models. They are ASCII files that can be copied and shared by other shell users.
3. SHOW - Displays the map resulting from a query.
4. PLOT - A set of routines for creating plot files and printing single maps or batch plotting and printing maps the user has put in a queue.
5. REPORT - Creates a default report or user defined report of attribute data and prints it to the screen, a file, or printer.
6. BUFFER - Routines for creating buffers on points, roads, bodies of water, and forest stand boundarys. Buffers can be combined and evaluated against previously created maps.
7. UPDATES - Allows the user to define and use 4 attributes associated with each stand so a user can flag a stand for a specific condition.
8. TWP FILES - A set of data loading and file management routines. Also contains a routine for creating a subset of township data.
9. EPPL7 - Provides direct access to the full featured EPPL7 GIS for users who need to do processing routines not included in the shell.
10. LEAVE - Exits to DOS.

(Phase II EPPL Shell - Users Guide, 1991)

Throughout the shell the user is always led by menus or screen prompts. Help screens are available for all menu items. Menus are intuitive and easy to use. Processing speed is very good, with the more complex queries taking up to 3 minutes to process on a 286 and under 2 minutes on a 386. Creating a plot file for printing a map can take up to 30 minutes. Running a query on many townships in a queue will naturally take longer also. Both processes can be put into a user defined queue and batch processed overnight or anytime the PC is not being used.

The average field manager will be productive on the shell with two to three hours of training. The user does not need to know GIS, although a basic conceptual knowledge is helpful. Far more important is a through understanding of the inventory data, its shortcomings and potential uses. DOF experience is that imaginative users will recognize uses of the data that they had never envisioned and become very productive with the shell. The shell allows the user to concentrate on the data and its use, not on learning a software application.

## Working smarter with the phase II EPPL shell

The DOF is implementing the Phase II EPPL Shell in field locations where users request it. Since the shell is basically a stand-alone analytic tool, not limited to any one management program, users are free to apply it as they see fit. The uses so far have been innovative and productive. The rest of this paper will focus on three uses that are allowing the DOF to "do more with less" and "work smarter."

### Overcoming limitations of data or database structure

The Phase II inventory was not designed as a digital database. However, using the EPPL shell as an analytic tool can provide information that would be nearly impossible to get from the tabular data alone. An example is a project to identify Old Growth stands.

Identification of Old Growth stands was a use that the developers of the Phase II inventory never envisioned. Old Growth wasn't an issue in Minnesota in the late 1970's and crucial data helpful to identifying old growth wasn't included in the database design. Needless to say, it is an issue that resource management agencies must deal with now. In 1990 The DNR started a project to identify and manage Old Growth stands on state land. The DOF's part of the project was to identify potential Old Growth stands and do a preliminary assessment of their suitability as Old Growth.

First field managers did a query of the Phase II stand attribute data using criteria developed by an inter-disciplinary team. The selection criteria were phrased in terms of Phase II stand attributes because that was the best data available. For example, the criteria for lowland hardwood forest types is:

1. Ownership = state owned.
2. Cover type = Lowland Hardwood.

3. Average tree diameter at breast height (DBH) >= 10 inches or average age >= 120 years.
4. Stand size >= 20 acres.

Field foresters then made an initial suitability assessment based on their personal knowledge of the stands in their area. Finally, plant ecologists from the DNR Natural Heritage program make field inspections of the revised list of stands and make a final determination of a stand's suitability as old growth.

The old growth selection process had already been completed before the EPPL shell was available. However, we found that we could have done a better analysis of the data if we had used the shell because it overcomes a basic design limitation of the inventory. In the Phase II inventory any feature that is 5 acres or larger is classified as a separate stand. Therefore, a stream, if it is greater than 5 acres, can be classified as a stand. A stream classified as stand can divide a forest stand into 2 separate stands though they actually belong to a uniform forest type. The result is 3 stands; 2 identical forest stands and a stand that is a stream.

Streams are common in lowland hardwood stands and sometimes divide a stand larger than 20 acres into 2 or more stands each less than 20 acres. A query of the stand attribute data using the lowland hardwood criteria would not show stands less than 20 acres as potential old growth stands. Doing the same query with the EPPL shell and not including the size limitation immediately showed when such a situation had occurred. The limitation in inventory design was overcome and a more realistic list of old growth candidates stands could have been identified. Also, a map of the stands, with road overlays, could have been made for the ecologists to use to plan field visits.

### Supplementing field level management decision making

The DOF uses an application called Timber Management Planning Information System (TMPIS) to make stand-level management prescriptions. The process uses an inter-disciplinary (primarily forestry and wildlife) team to identify stands needing harvest, regeneration, and sivicultural or salvage treatment for the next 10 year period.

First the team develops the basic criteria to identify the stands they want to review for the above listed stand treatments. Then the TMPIS program displays all the attribute data for each stand that meets the criteria. The team reviews the data and decides on a stand by stand basis what management treatment the stands should have. They can use the paper cover type maps to look at spatial relationships to supplement the attribute data. However, using the maps is a cumbersome, manual process and only shows the cover type, cover size and cover density relationships. The maps do not provide and site quality or age class information for example.

The DOF's Warroad field office is using the EPPL shell to improve the TMPIS decision making process. The standard TMPIS is run on one PC and the EPPL shell on another. If there is any question about the prescription for a stand, a query can quickly be developed on the EPPL shell PC and the spatial relationships in question examined. For example, the forester wants to harvest a stand of mature aspen but the wildlife manager believes that the harvest will result in too much immature aspen for best Ruffed Grouse management. In a few minutes a query of aspen by age class will generate a map that shows the aspen age classes and their spatial relationship. The management decision can be based on all relevant data as opposed to being an argumentive, subjective decision.

In addition, repetitive questions can be modeled using query logic files and batch processed on all townships in the TMPIS project area. The resulting maps can quickly be displayed whenever those questions come up. For example, timber harvesting next to White Tailed Deer winter yard areas may have a negative effect on the deer yards. A deer yard model can be developed in a query logic file that can be run on all townships in batch mode before the TMPIS process starts. The maps showing the location of deer yards can be displayed whenever the management decision needs that information.

The DOF is monitoring the Warroad TMPIS project to gather information on how much the EPPL shell adds to the process. The DOF is starting to plan the next generation of TMPIS and many believe it should be a GIS application. Warroad's experience will be very helpful in deciding if and how GIS should be incorporated into the next version of TMPIS.

### Using the EPPL shell to do field work more efficiently

Another use for the Phase II EPPL Shell was developed in the Littlefork field office. It clearly shows how the shell can be used to "work smarter." The Littlefork field office used the EPPL shell to create maps that show all their field work needs. Field staff use the maps to plan their field work more efficiently.

Many of the management activities and associated business functions done by the DNR are defined as distinct "programs." Consequently, field managers tend to think in terms of a specific program as they do their field work. For instance, foresters may appraise timber for the timber sale program one day, check plantation survival rates for the silvicultural program another day, and do field inventory work for the inventory program a third day. Often the stands needing work are adjacent to each other and the field work could be accomplished in 1 field trip instead of 2 if the field manager had been concentrating on "field work" instead of "programs."

To make the maps, the Littlefork forestry staff identified the field work involved for each program they are responsible for. They also identified "sensitive areas" that might affect management activity in a stand. This is the resulting list.

1. Plantation survival surveys.
2. Timber appraisals.
3. Post-harvest regeneration surveys.
4. Reinventory survey.
5. Old Growth stands.
6. Scientific and Natural Areas.
7. Recreation sites.
8. Heron rookeries.
9. Eagle nests.
10. Deer wintering yards.

These lists are entered into the database via the UPDATE and QUERY LOGIC file functions. For instance, a deer yard is modeled in a query logic file and stands having eagle nests are flagged using the UPDATE function. The maps show all field work needs and their spatial relationships to sensitive areas. The foresters will use the maps over the next year to plan field work based on total work needs in an area independent of program definitions. The following scenario shows how the maps can improve field work efficiency.

## Pre-EPPL shell

A forester is preparing a timber sale. He/she drives to the site for 1/2 hour, walks into the site for 1 hour and appraises the stand for 2 hours and returns the office.

## Post-EPPL shell

The forester checks the timber appraisal list and picks a stand that needs appraising. He/she then checks the work plan map and sees that a plantation survival survey needs to be done in a stand next to the appraisal stand. He/she drives to the site for 1/2 hour, walks into the site for 1 hour and appraises the stand for 2 hours. He/she then enters the adjacent stand and does a plantation survival survey for 1 hour and returns to the office. In both cases the total driving and walking time is 3 hours. If the assignments were made on a program basis, the tasks could have been done on 2 different days and/or by 2 or more people. The result would have been a total driving and walking time of 6 hours.

The above situation is hypothetical, but the potential for improved work efficiency is obvious if field work is viewed spatially and done independent of program definitions. The EPPL shell gives us the opportunity to capture those work efficiencies and to "work smarter."

## Conclusion

The Minnesota Department of Natural Resources - Division of Forestry has had to respond to decreasing budgets and increased demands on the resource. The DOF has found that GIS, despite its high initial cost in hardware, software, database development and user training, has proven to be a powerful tool to enable us to meet those demands. However, these costs have prevented its use by the field resource manager to support field-level decision making. The Division of Forestry is overcoming that barrier and providing GIS capability to the field manager with the Phase II EPPL Shell.

The Phase II EPPL Shell is a menu-driven interface to the EPPL7 GIS. It allows the field resource manager to access and query data and perform specified GIS analysis of forest inventory data and related management prescriptions. The shell is very easy to use and requires only 2 to 3 hours of training for the average field user. The user doesn't have to know GIS. Instead, the emphasis is on knowing the inventory data, its limitations and potential uses.

The most potentially productive use so far has been the development of field work planning maps that show the location of field work independent of program definitions. Other positive uses of the shell have been to supplement a stand management decision making process and providing better information by overcoming limitations in the inventory design.

Managers and field users alike are pleased with the ease of use and productivity gains offered by the Phase II EPPL Shell. We are confident that the use of this field resource management tool will enable us to continue to "do more with less" and "work smarter."

## Literature cited

Minnesota Department of Natural Resources. 1991. *Directions for Natural Resources: Challenges for the Decade.* Minnesota Dept. of Natural Resources, St. Paul, MN. 26 p.

*Forest Statistics For Minnesota.* 1990 (Unit Reports). USDA, Forest Service - North Central Forest Experiment Station, St. Paul, MN.

*Minnesota Forest Industries Factbook.* 1988. Minnesota Forest Industries Inc., Duluth, MN.

*Minnesota Forest Resources.* 1991. Minnesota Dept. of Natural Resources - Division of Forestry, St. Paul, MN.

EPPL7 Version 7 Release 2.0 - *USER'S GUIDE.* 1990. Minnesota Department of Administration, Land Management Information Center, St. Paul, MN 166 p.

*Phase II EPPL Shell - Users Guide,* 1991. Minnesota Department of Natural Resources - Division of Forestry, St. Paul, MN. 55 p.

# D3 Evolving methodologies

**A corporate database approach to GIS: techniques for integrating multiple GIS applications**

John Herring et al., Intergraph Corporation, Huntsville, AL, USA

**Resource management in the 1990s using statistical and spatial modelling**

Stephen Sperry et al., ERDAS, Inc., Atlanta, GA, USA

**Integrating spatial analysis and GIS through application program interfaces (API)**

James Strapp et al., University of Victoria, Victoria, BC, Canada

**Inexact spatial reasoning**

David Altman, Curtin University of Technology, Perth, Western Australia

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# A corporate database approach to GIS: Techniques for integrating multiple GIS applications

**Dr. John R. Herring**
**Kevin G. Pammett**

GIS/Mapping and energy applications,
Intergraph Corporation
Huntsville, AL
USA 35894-0001

## Abstract

*The greatest challenge for GIS applications in the 1990's is the integration of GIS technology into corporate data environments, characterized by large numbers of users with diverse backgrounds, and distributed heterogeneous database stores and applications. For this challenge to be met, two things must change in the way GIS does business.*

*First, GIS applications must integrate with one another and with other data applications and databases, such as word processors, presentation graphics, spread sheets, and statistical analysis packages. Data must be shared between software suites. Object-based feature analysis must be integrated with more traditional relational databases. Grid-based GIS must be integrated with vector-based topological solutions.*

*Second, GIS user interfaces must improve to make the underlying functionality more accessible. The value of a tool is in proportion to the number of people who can use it effectively. Today's GIS user interfaces require in-depth knowledge of the underlying software. The future GIS must be as easily learned as today's spread sheet. As difficult at these goals may seem, many of the tools needed to achieve them are here today. Advances in intelligent user interfaces, data modeling, and cognitive psychology are being applied to these problems and promise GIS tools that will allow everyone to "work smarter."*

## Introduction

Most forecasts into the future rely heavily on the advance of hardware, increases in processing speed and data transfer rates; or extrapolation on the adoption and usage by the GIS community of existing and proven software types, such as Relational Data Bases (RDB's) or Object Oriented Programming Languages (OOPL's) and Data Bases (OODB's), or the adoption of recent advances in GIS research. See Frank, 1991a, for GIS trends, and Marcus, 1991, for user interface trends. In this paper, we will look at a much more exciting and confusing development, the changes in system software brought about by two basic changes in the way data and processing are perceived in an integrated environment.

First, interoperability, the ability for multiple heterogeneous processes to share data, is rapidly moving from buzz-word to standards specification. Current GIS exchange standards, such as SDTS (Spatial Data Transfer Standard), DLG from the US Geological Survey, and DIGEST from NATO, emphasize the exchange of data at a macrocosmic level. Such transfers consist of "data files" or "geocells," making for complex data sets covering a significant area, and containing large amounts of data. Interoperability has a different type of data transfer requirement. Data in simultaneously running systems must be exchanged at an almost atomic level, feature by feature.

Second, the von Neuman concept of processing as a single stream of executing commands, is giving way to "non-von" processes, where multiple, independent processes share data, and cooperate to perform tasks in an asynchronous manner in a heterogeneous environment.

Both of these developments have made us rethink our methods of building applications.

First, we must think of the data structures as dynamically defined by the user to meet his needs at a particular instance. Static data structures, schemas, and data contents, are giving way to dynamically defined data views. Particularily relevant is the remote procedure call (RPC) standard that is embodied in the OSF DCE. This standard allows an application implementor to invoke procedures (subroutines) without foreknowledge of where the procedure might execute. Once this technology becomes pervasive, it will be possible to design distributed applications where the level of sharing is no longer just at the data or at the module level, but will instead be at the procedure call level. This opens up a whole new level of capabilities — like having standard bricks with which to build things, rather than having to hand-craft different things made out of non-standard "bricks" that cannot be used effectively together.

Second, we must learn how to build using the functionality provided by other applications. In an integrated environment where the complete functionality of many applications is readily available, there is absolutely no excuse for "reinventing the wheel."

And, third, we must integrate the applications at a user interface level so that the user's view of the system is seamless in the sense that he cannot, in general, perceive the transitions between applications. In fact, what the user perceives as a single command might use functions from many different applications. Without this consistent user view, we lose the predictability required for a good user interface. See Marcus, 1991.

Much is being said, written, and promised for the future. For example, using the terms from the Open System Foundation (OSF), the Distributed Computing Environment (DCE), the Distributed Management Environment (DME), and the Architecture-Neutral Distribution Format (ANDF) promise to allow for the sharing of data and functionality across systems which are heterogeneous both in hardware and software. Closely connected with this is the added dimension in data structuring that we get when the object oriented paradigm is introduced. Once objects or collections of objects are defined, the applications designer or database administrator has the capability of including arbitrary semantics into the scheme, not just a variation on the structure of the data. What we will address in this paper are techniques available and in use today to begin building applications that will eventually fit into the corporate environment.

The new paradigm thus requires applications to share both data and user interfaces. In the enterprise system, the user sits at the apex, controlling a broad range of functionality spread out across many diverse applications, through a relatively small and highly consistent user interface. Below this central applications bulge, the data and control flows converge on a global, distributed data base at the nadir of the system. In this paper, we will investigate some techniques for dealing with the two halves of this double pyramid. First, starting at the data foundations of our system, we will explore how some applications communicate. Afterwards, we will see how this data paradigm affects the way we must build the user interface apex of an enterprise system.

## The enterprise data paradigm

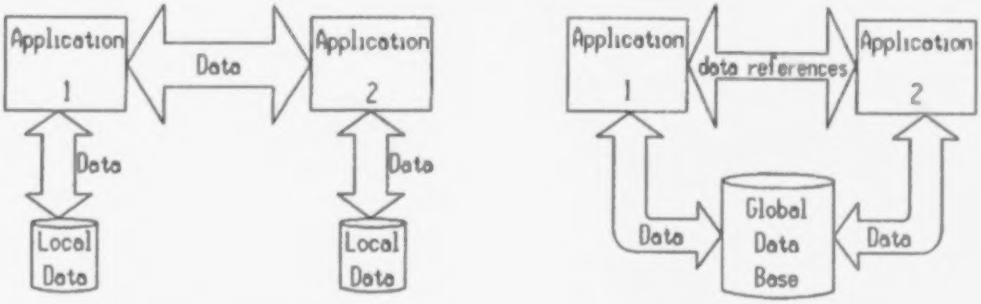
Each process can transfer data to any other process at run time, via a standard format for that particular type of data. This mechanism can move textual, and graphic data around quite nicely, using known and common standards.

The problem arises in GIS data when we try to use the standards from other applications, for example, a graphics or CAD standard. Most "graphic data" is just that, graphics — pictures without geographic references to tie the data to the ground or to attribute descriptions — and of little use to a GIS. When your are throwing simple pictures around, you can play fast and loose with the scale and make it "fit" the target area. GIS data is often more sensitive. If the target process cannot handle a particular type of data, a particular coordinate system, or a particular data model, the data transfers to such a process require a structural, coordinate or model transformation to a type, projection, or model understood by the receiving process. See Herring, 1991b.

As one might suspect from this, the hardest problem to address in building an enterprise environment is the simple fact that different applications hold strikingly different views of the data. For example, the old problem of integrating grid-based and vector-based GIS (Worboys,1991), or the controversy on object oriented versus relational forms. Standards usually do not help, since the applications often do not share data or processing paradigms, a prerequisite for any standard.

Another problem is the granularity of the data. Most standards reflect a "coverage" layered and horizontally tiled view for the partitioning of GIS data, where the smallest unit of data transfer is the map, data set, or layer coverage. In an integrated environment, the data movement between running applications may best be accomplished at a much finer granularity, say individual feature.

**Figure 1:** The two data sharing paradigms



Application Centered

Data Centered

### The data sharing models

There are two basic conceptual models for the sharing of data between applications: application centered, or data centered (see Figure 1).

In the application centered model, each application has a separate, independent, local data store. When they share data, each application is responsible for translating to one of the common formats, appropriate for transfer to the specific target application. This means multiple sets of translators, as many as one for each format and application. It also means a user interface problem, since it will often fall upon the user to choose the transfer parameters, including target application and transfer formats, an unnecessary level of detail for which the user is often either unprepared or indifferent.

In the data centered model, a global, shared database is at the center and the various applications communicate between each other by sharing the database. This means that the applications are responsible only for translating data from the views of the database to their internal formats. Given a database system capable of handling the multiplicity of GIS views of data, this translation requirement can be transferred from the applications to the database server. Applications can still speak to one another directly, but they transfer data by storing it in the global database, and signaling the target application with the information (usually a set of key attribute values for the target data items) needed for retrieval. In a hybrid model, both transfer methods are used, depending upon the circumstances. We can still classify systems as being one or the other model based upon the manner in which the "real" data is transferred. In the case of a GIS, "real" means geographic feature data.

In Intergraph's MGE world, the central geographic feature database uses an extended geo-relational model, with additional feature tables which communicate semantic information. This approach is very similar to the Spatial Data Transfer Format's (SDTS) use of data profiles.

### The schema problem

To establish the semantic equivalence between a geo-relational model and schema, and an object model and schema, we have to step back from the systems, look at where they came from, and note where they diverged. To do this we must look at data modeling. See Herring, 1991b. Theories vary as to names and assignment of labels, but the basic cycle is always the same; a conceptual model, as defined by a layer in the design hierarchy, is transformed into one more closely related to the physical implementation, usually one layer down in the data design. If this transformation step between one layer's model and the next layer's model has been automated, then we need not really deal with it. For example, the Relational Data Base data model of tables and query operations is mapped by each RDB vendor into physical storage and function libraries in different manners, the details of which are not important to any application written to take advantage of the SQL interface. The trick then is to make all of these "layer" transformations "transparent."

Thus, to begin, we need to reconstruct the "original" conceptual model, as this model has the greatest semantic content, a direct relationship with the way the user perceives the real world. The need for such semantic information was seen early on in the development of the relational model, and some work has been done in including it directly in the relational model's schema. See Codd, 1979, 1990. For us that is the entity, attribute, relation model from which the lower levels have been obtained. We then, for each application, devise an automatic sequence of data modeling tools that maps the common conceptual model to the physical realization for that application. Where older, existing databases are involved, the conceptual model must be "backward engineered" from the existing schema. If the design or the design methodology is well documented, this schema mapping step can be fully automated. This is most easily demonstrated through an example, showing how to integrate an object application into a classical geo-relational environment.

There are severe structural and semantic problems in heterogeneous environments, even when the paradigms for all the data handlers are equivalent. See Chatterjee, 1991. In a system involving both relational and object data systems, and both raster and vector based geometric feature representations, the problems are complicated by what appears to be irreconcilable differences in data storage and representation paradigms.

Intergraph has a unique position in that it has developed three products using three distinct data models for GIS; a Relational Data Base and linked geo-referenced geometry model (called geo-relational by Jack Dangermond), a grid based model (see Tomlin, 1990), and an object model (see Herring, 1987, 1988, 1989). The keystone products for these three models are MGE/SX, MGE/Grid Analyst (MGGA) and MGE/Dynamic Analyst (MGE/Dynamo). Even though these products use radically different data models, they are part of a single integrated GIS environment. We could spend a great deal of time on each of the processes, how they communicate with the database and how they use their special talents to advance the work flows, but let us concentrate on integrating the object oriented Dynamo into the geo-relational database environment.

The key to integration is the common, conceptual data model. Since relational and object models are incompatible at the physical level, we have to step up a level to the entity model from which they are both derived. This makes a great deal of sense, even between systems that "share the physical" model of data. The transition of conceptual models to physical realization within a relational database is not a precise mapping. Different design alternatives for data representation and different levels of normalization are possible.

However, looking at a relational schema, and knowing the manner in which it was designed, we can "backward engineer" to the root conceptual schema. From this model we can map easily onto the object schema within Dynamo. This may sound like a complex piece of legerdemain, but each of these steps is sufficiently well understood (see Wong, 1980; Bedard, 1989; Codd, 1979, 1990; Herring 1991b; Ullman 1988, 1989) that the schema translations can and are all automatic (as in the example in Figure 2). The MGE to Dynamo translators are "one button" push, and the input parameters are simply a list of features. The object schema is built on the fly during translation.

In the example shown in Figure 2, relational tables linked to geometric descriptions are mapped to features in the object space schema. Thus, the river table creates an object class called "river," and each geometric element with a linked "river" table row is translated as a "river" feature object in the Dynamo object space. Other relations are mapped either to complex features or feature to feature relations depending on the semantic information held in the "data dictionary" feature tables. Thus, the "river system" table builds complex features from the atomic line "river" features, and the atomic area "lake" features; while the "crossed by" table becomes a feature to feature relations linking each "bridge" feature (which can be either an atomic point feature or atomic "line" feature) to the "river" over which it crosses.

With this schema mapping established, Dynamo can exist in a data centered integrated environment with geo-relational MGE applications, reading directly from and writing directly to the RDB attribute database and the geo-referenced geometric elements.

### The data granularity problem

This data mapping approach leads us to a solution for the data granularity problem. Using a feature as the atomic entity, data transfers between the geo-relational and object applications can be accomplished at a very fine level of granularity. An analyst in Dynamo can transfer any number of features of any variety of types to or from the geo-relational global data base. Schemas in the object space and in the relational database are updated as the need arises to support particular transfers.

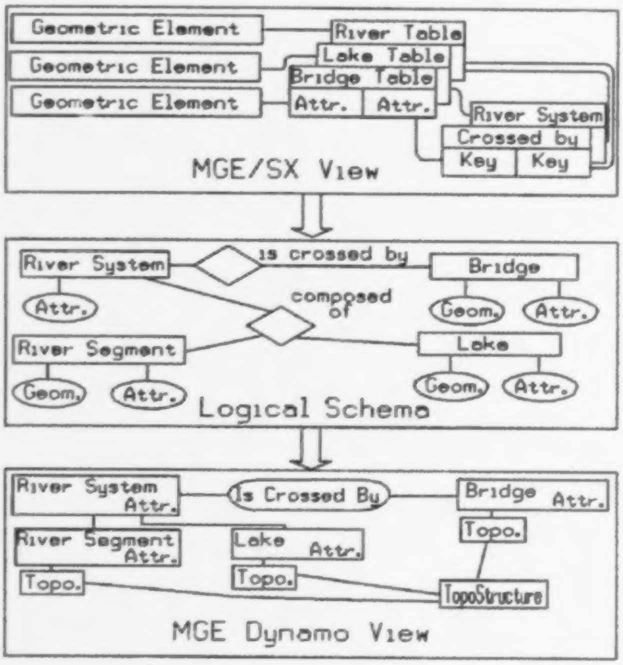
Once in the object domain, the data in Dynamo is still linked to the Relational Data Base, since the key attributes of each feature are maintained in the object data set. As the application runs, it can use its connection to the RDB to retrieve or to post new data based upon further feature lists or upon any analysis done within the object system. This allows us to use the analytic capability of the object system, with its stored procedures, integrated spatial query, dynamic segmentation, and dynamic data dictionary, while still maintaining the integration inherent in the overall geo-relational model. See Herring, 1987, 1988, 1989, 1990a, 1990b, 1991a, and Egenhofer, 1990.

This, in essence, puts a running Dynamo in an application centered environment with the RDB, using ANSI standard SQL interfaces mechanisms. Dynamo can create tables within the RDB, unload information derived either from the feature data, or from spatial and attribute analysis directly to those tables. This allows Dynamo to use the RDB's analysis, report generator, and applications environment. Furthermore, Dynamo is now linked in with the corporate data environment through the various RDB gateways, giving it access to the full range of applications capable of communicating with the RDB.

## The role of the user interface

The applications' user interface in an enterprise environment has three basic roles: integration, organization, and communication. Integration at the user interface level means that, as far as the user is concerned, there must be one and only one user interface. Organization means user

**Figure 2:** Example of schema migration from georelational to objects


Geometric Element
Geometric Element
Geometric Element
River Table
Lake Table
Bridge Table
Attr.
Attr.
River System
Crossed by
Key
Key
MGE/SX View
River System
is crossed by
Bridge
Attr.
composed of
Geom.
Attr.
River Segment
Lake
Geom.
Attr.
Geom.
Attr.
Logical Schema
River System Attr.
Is Crossed By
Bridge Attr.
Topo.
River Segment Attr.
Lake Attr.
Topo.
Topo.
TopoStructure
MGE Dynamo View

control of the interface layout; no matter how neatly or logically the designers of a user interface arrange their icons and widgets, it cannot be all things to all users for all tasks. Communication means that since the user is not going to read the vast encyclopedia of manuals associated with the large number of applications involved, unless he is in trouble, that the interface must be able to communicate to the user at any point in time his options and the meaning of the functions at his command.

**Integration of user interfaces: from many, one**

The integration problem is one of transition. Heretofore separate applications,will need to build transitions that appear to be seamless. In the common data environment, the application interface boundaries fade, and application boundaries are simply conveniences for the management of functions, and do not really represent different data. Since the user may use several applications to do his job, and he may in fact use them simultaneously, the applications must support a common "look and feel" in their interfaces. It should not be expected of the user to realize which application has the current control in input and output, and to remember the various peculiarities of the interfaces. In fact, there is no reason to expect the interface software to even be a part of the application. On the contrary, for a variety of reasons, the functional separation of functionality and interface into separate software modules is a very good idea. Different users of the same functions will require different interfaces, possibly due to application area, personal preferences, or reasons of nationality.

To accomplish this integrated interface, requires more that common button and icon software, it requires a common user interface metaphor. See Frank, 1991b, Kuhn, 1991. A metaphor is a correspondence of one cognitive domain to another, allowing for the user of knowledge and understanding from the first to be used in the second. For example, we all understand maps, and a GIS user interface built upon a map metaphor would use that understanding to give the user a set of expectations as to how the icons and widgets should work. The most common metaphors used are probably the camera and window manipulation, hence "zoom," and "pan," which arise from photography, and "frame" from the structure of windows, are applied to display interface. Hence, the interface metaphor imposes a familiar structure on unfamiliar situations, supplying the user with a consistent set of expectations.

### Organizing around the needs of the user

Any attempt at a universal user interface configuration is, at best, a compromise between the different needs of the various general types of users. The needs of the data analyst are different from those of the data capture specialist. Users of the same type will have different working patterns, and even the same user may change working style when confronted with an unusual data situation.

The sensible approach is to design tools that allow the user to modify the user interface to suit the needs of the moment. One such tool is the user configurable menu palette. Such a palette can be filled, through "click and drag" manipulations, with icons from the standard, application organized interface. These icons transmute to context-free versions to allow for the loss of the subpanel context. Palettes can be named, stored for future use, and distributed to other users. Several palettes can be active at one time, placed by the user at convenient positions on the screen. This allows the user to configure many icon palettes, and recall them in any combination as the task dictates.

### Communication with the occasional user

The corporate database's open philosophy presents a new problem for the GIS developer - the occasional user. Up until recently, the GIS community has been a closed club of professionals. People with GIS problems went to these "gurus" with their problems, which were usually grudgingly solved, accompanied with descriptions of how fortunate the requester was to have such a Guru available. This is changing rapidly. Dilettantes, needing answers to a few specific, and often highly technical, questions, now have access to massive amounts of geographic information, a great deal of it in the public domain, supplied by government agencies on CD/ROM.

The problem is that the GIS applications are not really ready for the casual user. This does not mean that the casual user cannot use a GIS,it means that most GIS's require him first to become an expert before many types of meaningful work can be done. What is worse is that expertise in one vendor's applications seldom carries over to another, so that if a particular function is missing or difficult to do, the casual user is stuck because he is incapable of accomplishing all the steps necessary to transform the data, and the question, to the form needed by another vendor's application.

Online help may answer some of these problems, but we must remember that an overly complex help file is no more useful than an overly complex manual. Appropriate choice of user metaphors, in addition to aiding the integration of diverse products, makes the interface more natural, hence simpler to use. Since"simpler" means that the user will have access to more of the functions, the metaphor enhances the power of the GIS by enhancing its usability GIS data browsers and "executive information systems" are being introduced, for example ArcView from ESRI. Although these products are yet to prove themselves over long term use, the understanding of metaphor and user-organized interfaces promise success for these systems.

## Summary

The two keys to building the integrated corporate data environment are flexibility and control. The data structures within an application must be flexible and dynamic enough to accept, retain, and supply data from or to a variety of other, possibly incompatible, applications. Most important, this data movement must be under sufficient semantic control so that the true meaning of the data is not lost or distorted through multiple data transfers. Different data models, and even different data modeling paradigms must be used in the applications within a single integrated system.

The key to using the integrated corporate data environment is the user interface. With metaphors to simplify and integrate the man-machine interface, and user organized interfaces to adapt it, GIS will move out of its current awkward "software for gurus" stage, into a truly user-friendly environment for use by anyone with a problem to solve.

### Trademarks

Intergraph is a registered trademark and MGE is a trademark of Intergraph Corporation. Other brands and product names are trademarks of their respective owners.

## References

Bedard, Yvan,, and Paqquette F. (1989) 'Extending Entity/Relationship Formalism for Spatial Information Systems', Proceedings of AutoCarto9, April 1989, Baltimore, Maryland.

Chatterjee, Abhirup and Arie Segev (1991), 'Data Manipulation in Heterogeneous Databases,' SIGMOD Record, Vol. 20, No. 4, December 1991, pp 64-68.

Codd, E. F. (1979) 'Extending the Database Relational Model to Capture More Meaning', ACM TODS 4, No. 4, December 1979.

Codd, E. F. (1990) The Relational Model for Database Management: Version 2, Addison-Wesley Publishing Company, Reading, Massachusetts.

Egenhofer, M., and Herring, John R. (1990), 'A Mathematical Framework for the Definition of Topological Relations', Proceedings of the 4th International Symposium on Spatial Data Handling, Zurich, Switzerland, Vol 2, pp 803-813.

Frank, Andrew U., Max Egenhofer, and Werner Kuhn, (1991a), 'A Perspective on GIS Technology in the Nineties,' Photogrammetric Engineering & Remote Sensing, Vol. 57, No. 11, November 1991, pp 1431-1436.

Frank, Andrew U., (1991b), 'Beyond Query Languages For Geographic Databases: Data Cubes and Maps,' International Workshop on DBMS for Geographic Applications, Capri, Italy, May, 1991.

Herring, John R. (1991a), 'Using Spline Functions to Represent Distributed Data', Proceedings of AutoCarto 10, ACSM, ASPRS, March 1991, Baltimore, Maryland, p46-58.

Herring, John R. (1991b) 'The Mathematical Modeling of Spatial and Non-Spatial Information in Geographic Information Systems' in Mark, D. and Frank, A., editors, The Cognitive and Linguistics Aspects of Geographic Space, Proceedings of a NATO ASI held in Las Navas del Marquis, Spain, July 1990, (to appear)

Herring, John R. (1990a), 'The Definition and Development of a Topological Spatial Data System,' in Kobel, O. editor, Photogrammetry and Land Information Systems, Press Polytechniques Romandes, Lausanne, Switzerland, pp. 57-70.

Herring, John R. (1990b), 'TIGRIS: A Data Model for an Object Oriented Geographic Information System,' to appear in Computers and Geoscience, issue on GIS Design Models and Functionality, presented 22 March 1990 at the Midlands Regional Research Laboratory, University of Leicester, UK.

Herring, John R., (1990) 'Using Category Theory to Model GIS Applications,' Proceedings of the 4th International Symposium on Spatial Data Handling, Zurich, Switzerland, Vol 2, 820-829.

Herring, John R. (1989),'A Fully Integrated Geographic Information System,' Proceedings of AutoCarto 9, April 1989, Baltimore, Maryland, pp 828-837.

Herring, John R., Larsen, R. C., and Shivakumar, J. (1988), 'Extensions of the SQL Query Language to Support Spatial Analysis in a Topological Data Base,' GIS/LIS'88 Proceedings, ACSM, ASP/RS, AAG, URISA, November, 1988, pp 741- 750.

Herring, John R. (1987), 'TIGRIS: Topologically Integrated Geographic Information System,' Proceedings of AutoCarto 8, March 1987, Baltimore, Maryland, pp. 282-291.

Kuhn, Werner (1991), 'Are Displays Maps or Views?', Proceedings of AutoCarto 10, ACSM, ASP/RS, March 1991, Baltimore, Maryland, pp 261-274.

Marcus, Aaron and Andries van Dam (1991), 'User-Interface Developments for the Nineties,' IEEE Computer, September, 1991, pp 49-57.

OSF ANDF (1991), OSF Architecture-Neutral Distribution Format Rationale, Open Software Foundation, Cambridge, Massachusetts, June, 1991.

OSF DCE (1990), OSF Distributed Computing Environment Rationale, Open Software Foundation, Cambridge, Massachusetts, May 14, 1990.

OSF DME (1991), OSF Distributed Management Environment Rationale, Open Software Foundation, Cambridge, Massachusetts, September, 1991.

Szymanski, William J. (1991), 'GIS Applications Software in an Open System Environment: How Standard is it?', GIS/LIS '91 Proceedings, ASP/RS, ACSM, AAG, URISA, AM/FM International, Atlanta, Georgia, October 1991, pp 418-426.

Tomlin, C. D. (1990) Geographic Information Systems and Cartographic Modeling, Prentice Hall, Englewood Cliff's, New Jersey, 1990.

Ullman, J. D. (1988) Principles of Database and Knowledge-Base Systems, Vol I, Computer Science Press, Rockville, Maryland.

Ullman, J. D. (1989) Principles of Database and Knowledge-Base Systems, Vol II, Computer Science Press, Rockville, Maryland.

Wong, Eugene and Randy H. Katz (1980), 'Logical Design and Schema Conversion for Relational and DBTG Databases' in P. Chen, ed., Entity-Relationship Approach to Systems Analysis and Design, North-Holland Publishing Company, Amsterdam, 1980, reprinted in Stonebreaker, M., ed., The INGRES Papers: Anatomy of a Relational Database System, Addison Wesley, Reading, Massachusetts, 1986, pp 395-408.

Worboys, M. F., and S. M. Deen (1991), 'Semantic Heterogeneity in Distributed Geographic Databases,' SIGMOD RECORD, Vol. 20, No. 4, December 1991, pp 30-34.

GIS'92

Symposium

Vancouver
British Columbia
February 1992

# Resource management in the 1990s using statistical and spatial modelling

**Lawrie E. Jordan, III**

**Stephen L. Sperry**

ERDAS, Inc.
2801 Buford Highway NE, Suite 300
Atlanta, Georgia, 30329 USA
Tel: 404/248-9000
Fax: 404/248-9400

**Abstract**

*Remote sensing and GIS have traditionally been separate raster processing programs with little or no integration. The image processing of satellite images has always relied on the statistical manipulation of digital data. Raster GIS started in the 1970s with primitive spatial operators and neighborhood capabilities. Boolean logic was not introduced to the raster GIS until much later. Despite this disparate background, users need and desire the integration of remote sensing and GIS to accomplish demanding applications tasks.*

*This paper will examine the directions of raster processing with an emphasis on statistical and spatial modeling. Key components are the use of statistical and spatial operators for GIS analysis and Boolean logic for image processing for feature extraction and land cover classification. The use of 32-bit real numbers instead of just 8 and 16-bit integers and the integration of ancillary data are all important in predictive models and as tools for expert systems capabilities. The result of the integration is a raster modeling language that not only replicates vector GIS capabilities, but also offers the tools for statistical modeling that are essential for the predictive models of the 1990s.*

## Introduction

The GIS industry experienced tremendous growth in the 1980s. Statistical and spatial modeling are becoming crucial functions of GIS. In the past, modeling could be done, but it was a manual process. Instead of computer scripts, a computer model required steps to be recorded in a notebook. Information layers such as proximity to population, surface hydrology, terrain features, and buffers could be merged and analyzed, but only after several preparatory steps. First, all files had to be the same size and resolution. Then intermediate files had to be saved. Once the files were uniform, data could be overlaid, aggregated, weighted, merged, indexed, etc., but each of these was a separate step. What was not possible was the generation of statistics. These statistics have become increasingly important to modeling and are now being incorporated into many GIS packages.

## Image processing is GIS

GIS has two major components: raster data and vector data tied together by a tabular data structure. Many people have thought of a raster image as a 'dumb' image, with just image statistics and no intelligence or attributes. A vector image has traditionally been seen as a 'smart' image with a topological structure and the ability to query tabular data. This limited view is rapidly changing. Now, both data formats are being linked to tabular databases with relational database capabilities.

Traditional image processing features many mathematical functions. These functions include ratios, tasseled cap transformations, and convolution filters. For example, Landsat Thematic Mapper data can be displayed in false color, so that information not visible in the original true color image can be inferred from this new representation. These data can be manipulated further with procedures such as principal components analysis, which calculates the major statistics in the image to extract wetlands and other features. A tasseled cap transformation will show still other features. Other standard image processing functions include the merging of Landsat TM data with SPOT Panchromatic data for a 10 meter multispectral product.

GIS and image processing have traditionally been divided, not only in functionality, but also on the types of files on which these techniques are used. Image processing functions have been reserved for multiband, continuous tone files. GIS functions have been restricted to single band, categorical files. It has been only recently that these distinctions have begun to blur. All along, there have been sophisticated users applying image processing functions to categorical files and GIS functions to continuous tone files. The trends are now leading toward incorporating statistical and spatial modeling into one package. Procedures that took hours and days in the past can now be accomplished in minutes with models. The ERDAS GIS modeling program GISMO combines image processing capabilities with Boolean logic and conditional statements for new decision rules using a scripts programming language.

When parameters were changed in the past, they were often difficult to isolate and time consuming to correct. Now, a model can be changed and run again in minutes. These models can be stored in libraries and used over and over on different data sets. This encourages a great deal of user interaction. Results are generated quickly, freeing analysts to concentrate on project objectives, rather than running the software.

In the future, all of the functions of image processing and raster GIS modeling must be combined for use on all file types. These functions should include spatial algorithms, convolution filtering and neighborhood analyses. Current trends are moving toward a graphical user interface (GUI) facilitating a short learning curve for new users. Ten years ago, the GIS and image processing market consisted of only scientific users. Today's users are experts in their fields with varying levels of geoprocessing experience. However, they possess a keen interest in learning and applying this technology. The users of tomorrow will have less and less technical training and increasingly diverse needs. They will require answers quickly and easily.

System aesthetics cannot be underestimated, but system functionality will still determine the leaders in the industry. Raster data, vector data and annotation must all be tied together as one layer linked to a tabular database. Traditional user-defined methods of image display such as standard deviation, covariance, correlation matrix and user-defined lookup tables will still be important. However, the emphasis will shift from simple display and enhancement, or simple GIS analysis, to the more comprehensive functions available in spatial and statistical modeling. Maximum likelihood classification is essentially statistical modeling. GIS analysis is really spatial attribute modeling. These functions must be linked. A new multiple data structure should include raster layers, histograms, lookup tables, descriptors and attributes. Users must be able to perform region of interest, algebraic, and logical operations on the same data set to compare the results of each.

GIS analysis has predominantly been thought of as a vector capability. One of the strengths of vector data is topology. Topology is basically a method of determining location and neighborhood relation in a study area. How does the terrain affect the analysis? This functionality must be applied to raster data. In ERDAS IMAGINE™, the latest release from ERDAS, Inc., image processing functions such as classification can be performed in a model. As conditions change and additional data become available, the model can be run again, keeping the information up-to-date at all times. An important application of this technology is the search for environmentally sensitive areas. Users will have the ability to identify up to 4 billion features with double precision accuracy. Image processing functions will include contrast stretching, histogram equalization, convolution filtering, principal components analysis and its inverse, Fast Fourier transformations, color transformations (IHS to RGB and vice versa), contiguity analysis, proximity analysis, neighborhood analyses such as boundary, standard deviation, maximum, minimum, majority, as well as topographic modeling with slope, aspect, relief and contour. The user-defined kernels for neighborhood analysis can be up to 255x255. This allows complex operations in a single model.

## Visualizing the future

The integration of statistical and spatial modeling will give users powerful tools for visualization. It is no longer enough to show simple contours and vectors. Today's analyses demand consideration of the effects of terrain, such as shaded relief. These realistic displays allow both lay people and decision makers to understand the results of complex models. Dynamically linked image windows provide multiple views of the same area with virtual roam and zoom, giving users access to more information quickly, using data with different resolutions and spectral characteristics.

These analyses should not be restricted to two dimensions — they should be three-dimensional. Three-dimensional analysis is an essential component of future geoprocessing systems. Users must have the ability to display, query and perform other analyses in the third dimension. They need to see true perspective views in a real-time mode. Animation in three dimensions will also be important. It will help users clearly visualize and understand the impact of an analysis.

## Data formats

As the technology is changing, so are the applications to which that technology is applied. With more applications there is a need for more types of data. Previously, 16-bit unsigned integer data was the standard. Now, 32-bit signed and unsigned integer data with real and complex values are available. This will allow greater dynamic range and precision of analysis. Satellite imagery, raster and vector data, demographic information, spreadsheet data, and many other data types that currently exist, or are being developed, need to be integrated quickly and easily. Once data are gathered, analysis begins, and the need for answers is immediate. Users need systems that are easy to learn and easy to use, products that are well tested and well supported, and products that handle that data they use and the new data they will need in the future. Data integration and transfer are essential for the industry to grow.

## Conclusion

The distinctions between raster and vector data will merge. Tomorrow's user will not be concerned with the data itself, but only how to use it. The handling of the data will be transparent. The computer will decide how to manipulate it. The industry is approaching the challenge the renowned industry expert, Roger Tomlinson, posed several years ago. He proposed that analysts not only use computers to duplicate what they have been been doing by hand, but use the computer for new ways of visualizing data and producing products that would not have been possible without computers.

To accomplish this goal, statistical and spatial operators must be used for GIS analysis. Boolean logic must be used in image processing for feature extraction and land cover classification. The predictive models and expert systems of the 1990's require a raster modeling language that integrates statistical and spatial modeling, 32-bit integer data with real numbers and ancillary data.

# Integrating spatial analysis and GIS through application program interfaces (API)

**J.D. Strapp and**
**C.P. Keller**

Department of Geography
University of Victoria
Victoria, B.C. V8W 3P5
Tel: 604-721-7333
Fax: 604-721-6216

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

## Abstract

*As the number of GIS applications and users continue to grow, the demand for more comprehensive and advanced spatial analysis capabilities increases. GIS users are exerting increased pressure on vendors to satisfy their needs to analyze existing data. However, developers have found it difficult to meet this demand through the present vendor-client relationship. Vendors have been largely unable to offer generic problem solving tools and users are unable to adapt current packages to their specific applications.*

*In a move to enable users to develop their own analytical modules, some vendors will soon be offering application program interfaces (API). An API for GIS consists of a library of routines for managing and displaying spatial data and associated attributes. These routines are called through and incorporated into programs written in standard programming languages, such as C or Fortran. With a proper API, programming expertise and knowledge of applicable solution techniques users can develop systems to help solve application specific problems.*

*Keywords: Application Program Interface, API, Spatial Analysis, Modelling, GIS*

## Introduction

Geographical information systems (GIS) can be defined as computer systems that manage the input, storage, manipulation, edit, analysis, display, and output of spatial data and associated attributes (Berry, 1986; Burrough, 1986). The bulk of the past three decades of research, development and application efforts in GIS have focused on inventory components, that is input, storage, manipulation and edit. This has paralleled tremendous efforts on behalf of government and corporations to build digital land related information inventories.

Since the latter part of the 1980s users have begun to express the desire to use GIS to conduct statistical and mathematical analyses of existing data. Users are putting increased pressure on GIS vendors to incorporate analytical needs within commercially available GIS. Consequently, spatial analysis and the visualization of results has moved to the forefront of the GIS research agenda in the 1990s. This is felt by many to have been essential to ensure the continued advancement of these information systems (see for example Densham and Goodchild, 1989; Fotheringham, 1991; Rhind, 1988).

In this paper, we address the integration of basic and advanced analysis within GIS through application program interfaces (API). An API is a toolkit for software development. Specifically, we will examine the potential of API to aid in the integration of analysis and GIS.

The paper commences with a clarification of what we refer to as analytical capabilities. Current impediments to unifying spatial analysis and GIS will then be addressed. The next section places API in context with other options to customize GIS. API and its benefits for GIS customization are discussed thereafter. Finally, the limitations of API are explored.

Currently, we are aware of two initiatives to produce an API for GIS, those developed from the TerraSoft and GENASYS software. Vendors have informed us that both are in beta stage, undergoing testing, and have yet to be released to the public. We have had the opportunity to conduct some research with the TerraSoft API and therefore will use it as a case study. Where appropriate, we refer to other software packages with which we are familiar.

## Analytical capabilities of GIS

We noted above that the GIS research agenda has been slow to focus on the development of analytical capabilities of GIS. This statement should be clarified by observing that we refer to non-cartographic analytical capabilities. The bulk of all commercially available GIS have developed cartographic analytical capabilities in their systems, capabilities perhaps best classified under spatial data edit, measurement, manipulation and query. We include in this group the following examples: conversion between raster and vector, the building of polygons and topology from raw line files, rubber sheeting, edge-matching, spurious polygon removal, measurement of shape, perimeter, area, volume, distance and direction, calculation of polygon centroids, buffer generation, topological overlay, and search by adjacency and proximity.

The GIS research agenda has yet to concentrate on analytical capabilities of GIS that concern statistical analysis of spatial data and associated attributes, spatial operations research, modelling and simulation. Statistical analysis includes parametric and non-parametric univariate, bivariate and multivariate statistics, and methods of ordination. Under spatial operations research, modelling and simulation we include all analytical aspects of operations research and management science where the problem definition has a spatial component. This includes, for example, the capability to conduct linear programming, goal programming, multiple criteria analysis, capacitated and uncapacitated network analysis, connectivity analysis, location/allocation analysis, nearest neighbour analysis, and simple and complex spatial correlation and autocorrelation analysis.

It must be recognized that even an introductory coverage of generic, let alone application specific, aspects of analytical capabilities listed in the above paragraph would require a number of textbooks. Commercially available software packages exist to solve all of the above generic statistical and operations research problems definitions. The challenge is to link these capabilities efficiently with GIS in a user-friendly and often application specific manner to support analysis of spatial data.

## Impediments to integrating spatial analysis and GIS

Following tremendous efforts in the 1970s and 1980s to build land related information inventories using GIS, users are increasingly expressing the desire to analyze existing data for research, planning, modelling and simulation. Although analyses tend to be application specific, GIS clients have been exerting pressure on vendors to satisfy their analytical needs within commercially available GIS. Given the specialized nature of individual analytical requirements, it has become increasingly difficult for vendors to respond to these diverse desires of their clients.

Difficulties have been encountered for a number of reasons. First, vendors tend not to have the specialized staff needed to understand fully and satisfy requests for highly application specific routines. Second, clients generally are unwilling to fund the research and development necessary to produce a new analytical component. They argue that the bulk of all development costs should be funded by the vendor since the vendor subsequently can sell the value added to their GIS to a broader client base. Vendors, on the other side, are understandably reluctant to spend the considerable research and development costs associated with developing new analytical components, especially if they are likely to be of interest to only a few clients (Fotheringham, 1991). Third, vendors have found it difficult to translate specific analytical applications into generic problem definitions that are marketable to a broad set of GIS users. Finally, vendors who attempt to satisfy analytical requirements for very diverse applications will find their systems cluttered with a puzzling collection of limited, specialized commands. Many of these commands will not be used by the majority of clients, but will confuse the user interface and exasperate users.

## Existing methods to integrate spatial analysis with GIS

As a result of the above impediments, identifying a feasible and affordable way of integrating analysis in GIS has become a GIS software development obstacle. Three dominant strategies are being explored. First, and as already noted above, vendors can integrate client specific analytical modules directly into their source code. Second, it is possible to develop explicit data exchange formats to transfer data between the GIS and commercially available software packages specializing in statistical analysis or operations research. Third, vendors can offer GIS users the ability to indirectly access data stored in the GIS, allowing users to customize their GIS to include application specific analytical code. The three strategies are discussed in more detail below.

### Integrating analytical modules directly in the vendor's source code

Early efforts to integrate application specific analysis in GIS focused on the addition of analytical code directly in the vendor's source code. Vendors pursuing this strategy encountered a number of problems. They found that efforts to incorporate application specific analysis in their source code consumed considerable resources of staff, time and finances, especially if the problem had first to be translated into a generic problem formulation. The resulting software often satisfied neither vendor nor client, while leaving the software cluttered with a puzzling collection of limited specialized commands[1].

### Development of explicit data exchange formats to commercially available software packages

A number of GIS vendors have opted to support statistical and mathematical analysis by building data exchange mechanisms between their system and commercially available statistical packages, operations research programs, or programmable database management systems. Methods of exchange tend to be via a direct link, or through import/export routines. The advantage of linking a GIS to commercial statistics or mathematics software is that users are ensured access to generic analytical problem definitions implemented in well tested software, where data manipulation and routines are accessible by a logical and consistent command structure. The disadvantage is that the exported information is usually limited to non-spatial attribute data; that is commercial analytical software products tend not to excel in handling spatial information. Also, access to solution procedures solving generic statistical or mathematical problem formulations does not necessarily guarantee resolution of a highly user-specific problem, and data, once analyzed, cannot readily be graphically displayed for visualization unless returned to the GIS environment.

### User-customized GIS

A third solution therefore is to allow the user to customize a GIS to suit application specific analytical needs. From the user's perspective, perhaps the simplest form of GIS customization is to commission existing GIS vendors to supply customized packages. We noted above that there are a number of fundamental problems associated with this strategy, notably issues of costs, lack of application expertise in the vendor's office, and vendor reluctance to commit to this type of work. It therefore is unlikely that GIS users can depend on existing GIS vendors to meet their divergent needs. What is required, therefore, is for the user to be able to integrate application specific analysis within a GIS without relying on the vendor's programming time. This implies that the user be allowed to modify or enhance the commercial GIS product.

At present, a number of GIS do support user customization of their user interface. By allowing changes to the colour, content and arrangement of menus, systems allow users to be able to remove unwanted commands, to rename commands to suit their application needs or language preferences, and to rearrange menus. The norm is for customized user-interfaces to be in text-based command or table form. However, user-interface customization recently has been broadened to include graphical interfaces by the "Genius" option of the GENASYS software.

Another customization option now frequently offered in GIS concerns the inclusion of scripts or macro chaining. Scripts enable users to concatenate existing commands to form simple batch type programs. These scripts can reduce the number of keystrokes required for repetitive operations, and allow users, for example, to run time consuming GIS applications overnight. Lotus 123's macro capability is the most common example of a script. GIS examples include the command mode of ARC/INFO, GENASYS and SPANS, among others.

More advanced versions of scripts evolve to a proprietary computer language, complete with variables, interface commands and flow control. Proprietary languages can perform powerful functions. However, programmers will always be limited by the GIS vendor's language. Users must first master the new proprietary computer language before creating advanced programs. As well, the language likely will not support all of the features needed for complex programs, such as arrays, memory allocation, file handling, sorting and seeking. Finally, scripts will always run relatively slowly because they are not compiled before execution, but interpreted.

Another option concerns user defined commands (UDC). UDCs enable programmers to create commands in the language of their choice. They are compiled into an executable file, resulting in much faster execution than interpreted programs. These programs are then added to the parent software and run as if they were initially included with the system. A UDC is selected through an addition to the parent programs menu, and executed in an operating system shell. Consequently, users of a UDC must have working copies of the parent package.

To write a UDC, programmers must have a basic knowledge of a computer language, a compiler, and information about the file structure, including spatial data, of the parent package. Access is given primarily to data, not the routines of the parent package. However, the program may allow limited access to graphical routines.

The above options, although supporting GIS customization, still do not allow users to incorporate application specific spatial analysis within a commercial GIS. To create truly functional customized analytical modules, users will need to have access to more advanced commands under customization, an issue API aims to address.

## Application program interfaces

API is able to overcome several of the limitations of the above methods by reducing the composition to its rudimentary elements. In essence, API treats GIS as a library of routines for managing and displaying spatial data and associated attributes. Through an API, routines can be accessed by a program to run GIS functions to interact with the user, and manage, display, import, and export spatial and attribute data. The application specific analysis must therefore be coded and compiled by the user in the non-GIS programming language. An API thus only consists of code. Software cannot be run without first coding and compiling a program. A related software package may be used in conjunction with an API, but it is not required.

A large assortment of APIs exist in non-GIS applications. CodeBase, the Paradox Engine, and dbVista are examples of database management APIs. Other APIs include the MicroSoft Windows Software Development Kit and the Optimization Subroutine Library from IBM.

The code for an API may come in several forms. Vendors most commonly provide code in a library. Data types are often initialized in include files. Also, object files to be linked with an application may be included.

The code is divided into functions that are called from a parent program created by the programmer. For example, to create a program to find the mean centre of a polygon using an API the following pseudo-code would be used:

```
1. Call an API routine to open a map.
2. Call an API routine to display the map.
3. Call an API routine to instruct the user to select a
   polygon with the pointing device and record his or
   her response.
4. Call an API routine to open the polygon selected
   from the polygonal database.
5. Run program to calculate the mean centre
      a. Initialize two variables to zero.
      b. Find the number of nodes in the polygon
         from the appropriate API variable. (This
         variable was initialized when the polygon
         was opened.)
      c. Loop through all nodes summing both X
         and Y coordinates read from the
         appropriate API variables (an element in
         an array of variables also initialized when
         the polygon was opened).
      d. Divide the X and Y sums by the number of
         nodes to determine the mean centre.
6. Call an API routine to display the centroid on the
   screen.
7. Call an API routine to close the map, when desired.
```

In the above, API routines are used for the steps in the program involving database management, user interaction, and graphical display. The programmer's own code performs the specific analytical procedures. It is important to grasp that, ideally, programmers writing the application specific analytical code should not have to understand the inner workings of the GIS, nor of each API routine used. They should merely need to understand how to use the host GIS in general, what land related inventory is stored in the host GIS, what function individual API routines perform, and the input and output capabilities of each API routine. They must also know the programming language supported by the API.

In reality, current versions of API make somewhat different assumptions. Pioneering the development and use of API capabilities in association with GIS is not without difficulties, some of which may be resolved through time.

An initial difficulty is that GIS vendors must develop in-house expertise in API. The bulk of code required for an API likely exists within the vendor's existing GIS; but not in a format suitable for distribution to clients. Code therefore must be rearranged and combined into highly structured routines, and must be accompanied by meaningful documentation. The quality and structure of the routines and their documentation will depend directly on the GIS vendor's research and development team's ability to look beyond internal programming requirements. It will require personnel that can place themselves in the position of the API user.

Documentation of the routines themselves is often completed while programming, but users will require additional instructions and examples of how the routines interact. This is where, based on our limited experience, most current APIs fail. Inclusion in the documentation of examples of how API can be used perhaps is the most effective way of educating application programmers about an API's routines. We recognize that the production of working examples is one of the most time consuming aspects of any software documentation, but we feel that the resulting benefits make this extra effort worthwhile.

Giving users the capability to customize their GIS analytical capabilities utilizing API makes a number of assumptions that merit examination. For a start, it assumes that the user has access to proficiency in statistical and mathematical details concerning their application specific analytical requirements, including knowledge about the relative advantages of solution algorithms. One should be able to assume that application specific GIS clients have access to solid expertise in methodological issues specific to their area of specialization. However, it is less likely that the users will have readily available access to detailed knowledge about solution algorithms. Independent consultants should be able to help the user here.

At this stage in API development, users must not only have access to algorithm proficiency, they also need access to reasonably proficient computer programmers. The programmer using API must have knowledge of the programming language in which the GIS and the API were written; knowledge which ought to be almost equal to that of the programmers who wrote the GIS. In creating their original GIS, vendors presumably used a programming language to its fullest. Most likely, the API subsequently developed by the same company is simply a result of repackaging and documenting that original code. Therefore, to understand currently available API routines and data structures, users must have an understanding of standard programming conventions and complex data types. For example, the variable to access the arc number in the topological database of the TerraSoft API is a C structure five levels deep. Failure to understand the meaning of the last sentence implies difficulties when setting out to program using that vendor's API.

## Discussion and conclusion

API has a tremendous potential to alter the way users approach GIS. With an API, GIS can be adapted to meet users' specific needs; not, as presently, visa versa. API will allow analytical capabilities of a GIS to be customized to be as large or small as required by specific users. Analysis via API may turn out to be a small add-on to an existing GIS, or may become a full fledged analytical package independent of the parent GIS. API will clearly give the user more flexibility and options in using GIS, however, it places the onus of developing analytical capabilities on the user, not the vendor.

Why bother with API if it requires programming skills equivalent to those required to build the GIS in the first place? The advantage of a combination of GIS and API is that the programmer does not have to be concerned with graphics programming and the programming problems associated with geo-referencing and cartographic manipulation. If users are going to be willing to invest the time and effort to customize their GIS by using an API, they should be able to produce routines to satisfy their own analytical requirements in the most efficient, effective and user-friendly manner, while maintaining all the benefits offered by commercially available GIS at present.

API at what cost? Once a GIS vendor produces an API, a decision must be made concerning pricing policy. The pioneering development of an API contains years of programming time and expense in a format that is very valuable to a competing GIS vendor. Reluctance to distribute this information cheaply is understandable. Until GIS APIs are offered by firms other than those marketing GIS, prices are likely to remain high.

One of the advantages of API is that a program developed using an API is the property of the API owner, not the developer of the GIS or the API. API users may not distribute parts of the API, but unless run-time royalties are stipulated upon purchase of the API, they may freely distribute executable files containing the API routines. They are free to charge for their creation, without further payment to the API developer.

In order to continue the development and success of API, there must be continued research. Research is required in three notable areas. First, the programming expertise required to use an API must be reduced. Developing analytical capabilities using today's APIs requires substantial learning and coding efforts. Higher level routines should be developed so that more can be accomplished with less code, without loss of flexibility. An object-oriented approach would be beneficial. Second, API users ideally should not be limited to the programming language used in the API. The IBM Optimization Subroutine Library, for example, supports several languages equally. Finally, improved documentation is required to enable users to learn an API much more readily. More detailed descriptions and extensive use of examples are important.

## References

Berry, J.K. (1986) "Learning computer-assisted map analysis," *Journal of Forestry* 39-43.

Burrough, P.A. (1986) *Principles of Geographical Information Systems for Land Resource Assessment* Oxford: Oxford University Press.

Densham, P.J. and M.F. Goodchild (1989) "Spatial decision support systems: a research agenda," *GIS/LIS '89: Proceedings of the Annual Conference and Exposition* Bethesda, MD.: American Society for Photogrammetric and Remote Sensing, 707-716.

Fotheringham, A.S. (1991) "GIS and spatial analysis: an NCGIA research initiative," *Environment and Planning A* 23: 1390-1391.

Rhind, D. (1988) "A GIS research agenda," *International Journal of Geographical Information Systems* 2: 23-28.

---

*[1] Exceptions include the shortest path problem which most vendors have been able to incorporate relatively smoothly.*

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Inexact spatial reasoning

**David Altman**

Curtin University of Technology,
School of Computing Science
GPO Box U 1987
Perth, W. Australia 6001

## Abstract

*This paper describes research in GIS spatial analysis subject to the presence of uncertainty. Two sources of uncertainty are considered: the uncertainty which exists within the spatial data; and the uncertainty in the constraints imposed on an analysis operation. The spatial analysis operations under consideration are the class of spatial relationships related to proximity. A point data structure which embodies uncertainty is defined. Next a heuristic for measuring distance to an inexact point is defined. From this base we show how GIS queries which contain elastic constraints can be evaluated for an area and a new database layer created which represents the degree of conformity of the area to the query.*

## Introduction

This paper describes research in GIS spatial analysis subject to the presence of uncertainty. Two sources of uncertainty are considered: the uncertainty which exists within the spatial database; and the uncertainty in the constraints imposed on an analysis operation.

The research is intended to satisfy a requirement to deal with spatial data in imprecise, vague, or qualitative form. The approach taken is firstly to develop data structures which embody uncertainty in the object being modeled, and secondly to develop analysis techniques which are able to account for the uncertainty in the data. These analysis techniques also allow the use of vague and qualitative parameters.

The approach is essentially a numerical one: the uncertainty is quantified and mathematical techniques used in order to reason about the spatial relationships. Fuzzy set theory is employed both as a representational device and as a reasoning methodology in this work.

This paper is structured into five main sections. The first three sections deal with background material: a review of uncertainty; uncertainty in the field of spatial information systems; and an overview of spatial relationships. The next section describes the main subject material of the paper. This comprises five subsections: a definition of a point data structure which embodies uncertainty in location; a definition of a distance metric; a process of inexact reasoning about proximity; a grid cell representation of spatial relationships; and a generalised approach to inexact spatial reasoning. The last section concludes the paper and gives an outline of future research directions.

## Uncertainty

Uncertainty is a phenomenon which we encounter daily and account for routinely in our decision making. Equally so, there is a need to deal with uncertainty when we apply computers to the solution of real world problems.

Uncertainty originates from a number of sources (Graham & Jones, 1988): incomplete information; errors or imprecision in data; data in vague or qualitative form. The reasons for the presence of uncertainty may also vary:

equipment limitations or faults; communication difficulties; the cost or feasibility of data acquisition; conflicting information.

Uncertainty is not always undesirable. When we communicate verbally, we often prefer to be imprecise. We talk about people who are "middle age" without much concern about the range of ages that represents, nor would we wish there to be such a precise definition. Similarly, we communicate distances and spatial relationships in "fuzzy" terms. My office is "about 5km." from the city and my home is "next to" a park. Such phrases still constitute viable information.

In using computer systems, we are faced with the choice of whether to ignore uncertainty or discard the information altogether; or alternately to deal with it separately and compute some sort of reliability or confidence factor to be assigned to any derived data or results.

The first option may be appropriate in the case of one-off errors; however if data costs are high, data is scarce, or if the uncertainty is only partial; there is justification for attempting to accomodate the uncertainty in some way. There are presently four major approaches for doing this: probability theory; fuzzy set theory; theory of evidence; and non-standard logics. We will briefly review the first two which are the most commonly adopted approaches.

### Probabilistic methods

In the empirical sense, these methods involve computing probability factors based on frequencies of occurrences of a vector $X = \{x_1, x_2, \ldots, x_n\}$ of $n$ observed values. If the observed values in $X$ are subject to random uncertainty (due to measurement error perhaps), the probability distribution can be assumed normal and the observed values used to estimate the statistical parameters of the distribution. Subsequent observed values can thus be assigned a probability of belonging to a particular decision class $G$, ie., the conditional probability: $pr(G|X)$. Bayes rule for decision analysis (James 1985) simply assigns $X$ to the class with the highest conditional probability. The Bayesian formula for inference (Halter & Dean 1971) provides a method for deriving $pr(G|X)$ (posterior probability - which may not be possible to evaluate) using $pr(X|G)$ (*a priori* probability - which can be evaluated).

A probability function can also be developed by classical or subjective means. The classical approach is applicable when, for instance, we are observing the toss of a die. Subjective probabilities are, on the other hand, more akin to betting rates. The subjective probability of an event $E$ occuring is the amount we would pay for a return of 1 unit if $E$ does occur. In both cases the Bayesian formula is still applicable.

### Fuzzy set theory and fuzzy logic

*Fuzzy set theory* is an extension of ordinary set theory in which set elements are allowed to have partial membership (the *degree* of membership) of the set. The degree of membership of a fuzzy set is a real number on [0,1], where 0 indicates no membership and 1 indicates complete membership. Thus a fuzzy set has elements of the form $\mu/x$ ; where $x$ is the element value and $\mu$ is its degree of membership.

*Fuzzy logic* is a branch of fuzzy set theory in which propositional statements assume partial truth values (or degrees of certainty) rather than being wholly either true or false. This can be interpreted as assigning meaning to propositions (in canonical form):

$$P \rightarrow X \text{ is } A$$

where $X$ is a vector of $n$ variables; $X = \{X_1, X_2, \ldots, X_n\}$ and $A$ is a *fuzzy predicate* (Zadeh 1988). Its purpose is to provide an elastic constraint on values of $X$ by defining a *possibility distribution* for $X$:

$$\Pi_X = A$$

The degree to which a value $u$ of $X$ satisfies the constraint $A$ is given by:

$$poss\{X=u\} = \mu_A(u)$$

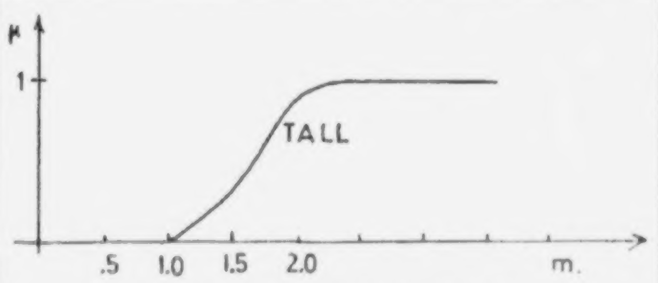


**Figure 1:** *The membership function for TALL.*

where $\mu_A$ is the membership function for $A$.

For example, if we wish to evaluate the proposition:

$$P \rightarrow HEIGHT(JOHN) \text{ is } TALL;$$

we see that $X = \{HEIGHT(JOHN)\}$ ; $A = TALL$ and if $u = 1.9m$ ; the possibility distribution for $TALL$ may reveal:

$$\mu_{TALL}(1.9m) = .9$$

which is the degree to which proposition $P$ is true. The membership function $\mu_{TALL}$ is shown in Figure 1.

Possibility and probability are sometimes confused but are entirely distinct concepts. Possibility is a measure of the *feasibility* or *conceivability* of an occurance. Probability is its *likelihood*. In a horse race it is conceivable that any horse could win, however, some horses are more likely to win than others.

With respect to uncertainty, the contribution of fuzzy set theory and fuzzy logic is the ability to partially classify an object; as opposed to the Bayesian approach which aims for conclusive classification by optimising the decision function and class separability. Furthermore, as the above example indicates, the theory provides a representation and reasoning tool for dealing with vague and qualitative data.

## Uncertainty in spatial information systems

Commercially available spatial information systems such as geographic information systems, image processing systems, pictorial database systems, and facilities management systems seldom acknowledge the presence of uncertainty in the data. There is usually no mechanism for observing the propogation of errors in data analysis operations. One of the research inititatives of the National Centre for Geographical Information and Analysis (NCGIA 1989) has focused on the need to be more concerned with the accuracy of spatial data and to develop data structures and functions which allow for error storage and propogation.

The fact that current systems adopt an *avoidance/dismissal* policy for dealing with uncertainty is reflected in architectural assumptions implicit in data models and query languages. We briefly review these in the following sections.

### Assumptions concerning the spatial extent and attributes

Current systems assume that the spatial extent of an entity can be crisply delineated, and that within that delineation, the attributes it possesses are uniformly concentrated. In reality, boundaries between classes are often indistinct, and spatial entities possess attributes in different concentrations, or to different degrees. For example, a GIS may represent a forest region bordering on a rice-growing area, whereas in reality there may be some inter-mixing of both classes about the common boundary.

### Assumptions concerning parameters input to analysis operations.

GIS analysis generally involves the formation of queries which are constrained in two ways (Peuquet, 1986):

(1) by the properties of individual objects; eg. area, length, centroid, shape, texture and other attributes;

(2) by the spatial relationships amongst the objects; eg. distance to other objects, nearest neighbour, overlap and containment.

These constraints form parameters which are input to analysis operations. They must be expressed in precise terms, for example, as a logical combination of attributes as in:

*height < 1000m AND rainfall > 100mm.;*

or, for a neighbourhood operation:

*within a radius of 5km*

Users of GIS, however, do not always wish to be so precise, or alternately, may not have the facilities or expertise to enable them to provide the precise parameters. With the spread of GIS and the down-sizing to less expensive workstation or PC based machines, there is a requirement to accomodate a wider range of users - some of whom will want to be able to view data relationships in an imprecise way.

In the example above, a rainfall of 99mm would probably contribute some meaning to the result; however it is omitted from the computation. It may have been intended to specify that the rainfall should be above *approximately* 100mm.

Fuzzy set theory can be useful in overcoming some of the above mentioned limitations because it allows us to consider spatial primitives with boundaries which are not crisply defined and which possess attributes which have partial class membership. Furthermore, fuzzy sets can been used as a means of numerically representing inexact quantities which appear in common speech. These are terms such as *YOUNG, OLD, ABOUT, QUITE, VERY, NEAR, FAR* and are refered to as *fuzzy linquistic variables*.

## Spatial relationships and reasoning

Spatial relationships can be classed as topological or metric. Topological relationships, such as *OVERLAPPING* and *CONTAINING*, are invariant under translation, rotation, and scale change. Metric relationships are measurement based. They can be related to direction; as in *ABOVE, BELOW, NORTH, WEST*; or distance; as in *NEAR, FAR, 5km AWAY*.

Directional metric relationships are invariant under translation and scale change but not rotation. Distance metric relationships are invariant under rotation and translation but not scale change.

A number of different approaches have been proposed for dealing with spatial relationships in spatial information systems. Typically, the data structure determines what sort of spatial relationships can be derived. Relational database systems, which have been used for GIS, can handle spatial relationships by explicit coding of adjacency information (Van Roessel 1987); but are constrained by a record oriented structure when performing spatial operations. Topological data structures store adjacency information explicitly; allowing efficient spatial searching. However, they are inefficient for evaluating arithmetic or logical combinations of the data attributes through overlay analysis. In cellular (or *tesselation*) data structures, such as a *raster (grid cell)* or *quadtree*, location is implicit in the data elements. Operations such as overlay are more efficiently performed with this structure.

Spatial reasoning uses known spatial relationships to derive new ones. Egenhofer (1991) has described eight binary topological relationships: *DISJOINT; MEET; EQUAL; INSIDE; COVEREDBY; CONTAINS; COVERS; OVERLAP;* in terms of the intersections of interior, boundary, and exterior regions of two polygons. These descriptions can be used in conjunctive phrases to derive new relationships according to a set of inference rules. Thus given:

*MEET(A,C) AND CONTAINS(B,C)*

we can derive some unknown relationship between *A* and *C*.

However, these spatial reasoning techniques do not incorporate uncertainty in the database, and are based on precise definitions of spatial relationships and constraints for reasoning. More recently, Dutta (1991) has presented a model for spatial reasoning in temporal and one dimensional space in which the events or objects have uncertain location. This paper builds on some of these concepts in two dimensional space.

## Inexact spatial reasoning

We now proceed to describe research which involves the use of fuzzy sets to model inexact spatial primitives and to analyse their spatial relationships. The spatial relationships which are being examined are the sort of imprecise qualititative relationships which a natural language interface to a GIS might possess. Such an interface would be expected to interpret elastic constraints such as *NEAR, FAR, ABOUT, BESIDE, SURROUNDING* and so on in order to define metric and topological relationships.

The use of such elastic constraints alongside other precise constraints would allow more natural and expressive queries such as:

"identify areas which are NEAR to a town, BESIDE a river and SUITABLE for agriculture;

where the spatial objects may themselves be imprecise in their delineation and attributes.

### Fuzzy points

We start with a spatial primitive which we shall call a *fuzzy point*, $P_f$

A fuzzy point is a fuzzy relation (the tuples are a fuzzy set) which maps points in $R^2$ (2D coordinate space) into possibilities on [0,1]. It is represented by an elliptical cone with base on the X-Y plane and height equal to one. The elliptical base can be located anywhere in the X-Y plane, with any orientation. This is shown in Figure 2.

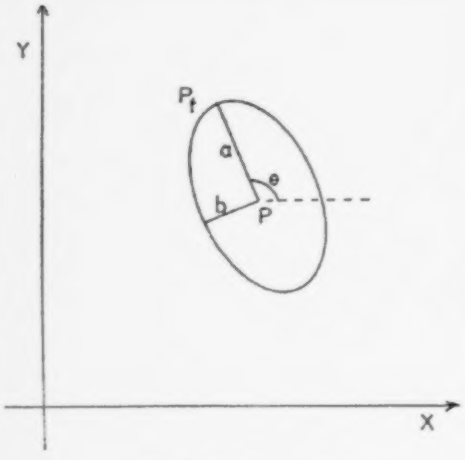


**Figure 2:** *A fuzzy point.*

The possibility for points in the X-Y plane which lie under the cone is computed by evaluation of the Z-ordinate of the equation for an elliptical cone whose base is translated and rotated. The centre point *P* lies under the vertex of the cone and hence has unit possibility; whilst on the perimeter of the elliptical base, the possibility value is zero.

It is convenient to think of $P_f$ as a point object whose exact location is unknown but which has the greatest possibility of being at *P*.

A fuzzy point $P_f$ is defined as a quintuple:

$$P_f \equiv (P_x, P_y, a, b, \Theta) \qquad \text{(D1)}$$

where:

$P_x, P_y$ = coordinate location of *P*, the centre of $P_f$;

$a, b$ = length of semi-major and semi-minor axes of elliptical base;

$\Theta$ = counter-clockwise rotation of elliptical base from semi-normalised position (major axis parallel to X axis).

A *normalised fuzzy point* is a fuzzy point which has been rotated and translated such that its centre is at the origin and its major axis is parallel to the X axis. Normalisation is performed to simplify computation.

Now, given a database of fuzzy points:

$$\varnothing = \{P_f, Q_f, R_f, \ldots\}$$

we can compute the possibility distribution for all points in $\varnothing$. This identifies the *spatial support* (possible location) of the points in $\varnothing$. An example of this is shown in Figure 3, where the brighter shades indicate higher possibility of location of the object.

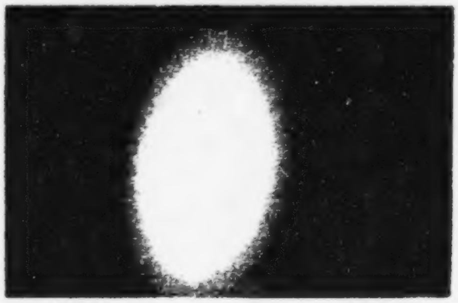


**Figure 3:** *Visualisation of fuzzy point.*

If we consider a single fuzzy point, $P_f$, its possibility value at a coordinate location $Q$ outside the spatial support of $P_f$ is zero. The possibility within the spatial support of $P_f$ is computed in two steps:

1. $P_f$ is normalised by translating $Q$ by $(-P_x, -P_y)$ and rotating $Q$ clockwise by $\Theta$ where $P_x, P_y$ and $\Theta$ are as defined in *(D1)* above.
2. The equation for a unit height elliptical cone translated by (0,0,1) is evaluated for the Z (possibility) axis. This simplifies to application of the formula:

$$\mu = 1 - \sqrt{(Q_x^2/a^2 + Q_y^2/b^2)}$$

where:

$Q = (Q_x, Q_y)$ *and* $\mu$ is the possibility.

### Fuzzy Distance

A distance metric is fundamental to an understanding of many spatial relationships. Thus, having defined the concept of a fuzzy point, the next task was to define the concept of distance between fuzzy points.

We firstly consider the one dimensional space where points are expressed as *triangular fuzzy numbers (TFNs)* (Kaufmann & Gupta 1988) which are triples:

$$T_f = (T_c, T_l, T_r)$$

*where:*

- $T_c$ is the location of the centre, and
- $T_l, T_r$ are the lengths of the left and right spreads respectively.

An example of a TFN is given in Figure 4.

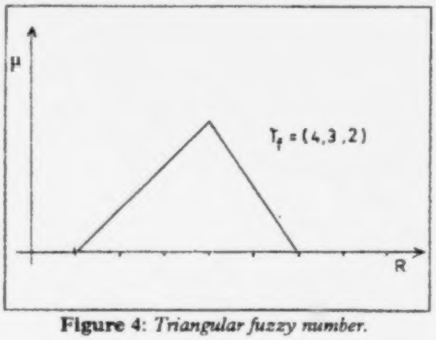


**Figure 4:** *Triangular fuzzy number.*

The distance between 2 points; $U_f, V_f$ in one dimension is simply their fuzzy arithmetic difference:

$$D_f = U_f -_f V_f$$

*where* $-_f$ is the fuzzy arithmetic difference operator. The result of this expression is a new TFN defined as:

$$D_f = (U_c - V_c,\ U_l + V_l,\ U_r + V_r)$$

This metric may be extended to the two dimensional case if we modify our original definition of a fuzzy point to become a triple:

$$P_f = (X_f, Y_f, \Theta)$$

where $X_f, Y_f$ are the fuzzy numbers $(P_x, a, a)$, $(P_y, b, b)$ respectively and $P_x, P_y, a, b, \Theta$ are as in *(D1)* above. In this case we can define a distance formula based on Euclidian distance:

$$D_f = \sqrt{}_f\,[(X_{2f} -_f X_{1f})^{2f} +_f (Y_{2f} -_f Y_{1f})^{2f}]$$

where the two points are given by the fuzzy numbers: $(X_{1f}, Y_{1f})$ and $(X_{2f}, Y_{2f})$ (the rotational angle $\Theta$ of both points is assumed zero for the moment). The operators: $2f$ and $\sqrt{}_f$ are fuzzy square and square root respectively and can be derived from fuzzy number multiplication $*_f$. Fuzzy number arithmetic operators are defined in Kaufmann & Gupta (1988).

Whilst this metric is simple to use, it means that the axes are orthogonal to each other with respect to uncertainty. That is, the uncertainty along the X (Y) axis due to $X_f$ ($Y_f$) does not contribute to that along the Y (X) axis; whereas off axis, the uncertainty is determined by both $X_f$ and $Y_f$. This creates an increase in the uncertainty as reflected in the spreads of the computed distance as the angular distance approaches multiples of $\pi/4$ and a decrease as it approaches the axes. This is an undesirable effect as we would wish the uncertainty to be independant of angular distance. For example, if we traced a concentric circle around a circular shaped fuzzy point we would expect that the uncertainty in the distance to the point remained constant.

Consequently, we have defined a heuristic for distance as described below. At present this is defined for the distance between an exact point (a fuzzy point with zero length axes) and a fuzzy point; future work will aim to remove this restriction.

The distance from an exact point $Q$ in $R^2$ to a fuzzy point $P_f$ in $R^2$:

$D_f = DISTANCE(Q,P_f)$

is expressed as a TFN in which the centre is the *most possible* distance and the spreads are determined by the range of possible distances.

It is defined as a triple:

$D_f \equiv (D_c, D_c - D_s, D_l - D_c)$

where:

$D_c =$ the Euclidean distance from $Q$ to the centre of $P_f$

$D_s =$ the "shortest" distance - ie. the length of the shorter normal from $Q$ to the perimiter of $P_f$

$D_l =$ the "longest" distance - ie. the length of the longer normal from $Q$ to the perimiter of $P_f$

The lengths $D_c, D_s, D_l$ are shown in Figure 5.

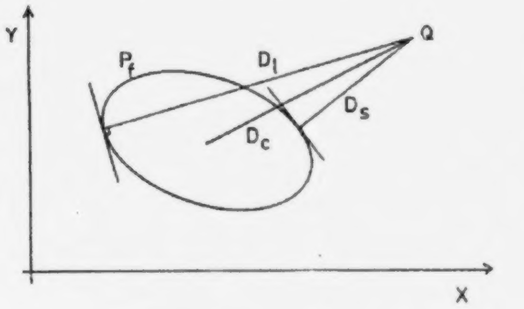


**Figure 5:** *Fuzzy distance.*

After normalisation of $P_f$, the lengths of the normals are determined by the solution of a biquadratic:

$f(\propto) = aQ_x sin(\propto) - bQ_y cos(\propto) + (b^2 - a^2) sin(\propto) cos(\propto)$
(D2)

where:

$\propto$ gives the angle of the diameter which intersects the normal on the perimiter of the ellipse. The point of intersection is used to compute the distance to $Q$.

$Q_x, Q_y$ are coordinates of $Q$, and

$a,b$ are the elliptical base semi-axes defined in *(D1)* above.

Numerical methods must be used to solve *(D2)*. Normally there are two roots, however four solutions arise when $Q$ is interior to $P_f$

**Fuzzy proximity**

With a distance function suitably defined, we are able to examine some spatial relationships which depend on distance. The present focus is on proximity, although it is intended to examine other classes of relationships.

Two basic fuzzy predicates for proximity are *NEAR* and *FAR*. These are considered here to represent a *unary* relationship. That is, we are considering the state of *being NEAR* to an object, rather than a relationship between two objects. This means the definition of *NEAR* is dependant on the object which it predicates. Hence if *A* and *B* are different objects the definitions of *NEAR(A)* and *NEAR(B)* may also be different.

We can define *NEAR* and *FAR* in terms of TFNs. The predicate *NEAR* can be defined as:

$NEAR(\ ) \equiv (0,0,T)$

*where* $T$ is the minimum distance at which an object has zero possibility of being *NEAR*.

If we now measure the distance from an exact point $Q$ in $R^2$ to a fuzzy point $P_f$, we can determine whether $Q$ is *NEAR* to $P_f$ by evaluating the truth value of the proposition:

$S \rightarrow Q$ *is* $NEAR(P_f)$

We can evaluate this proposition by computing the *composition* (Klir & Folger 1988) of the fuzzy relations $NEAR(P_f)$ and $DISTANCE(Q,P_f)$.

Composition is a means of forming a new fuzzy relation from two others which have one or more common domains. If we consider two binary fuzzy relations; $T(X,Y)$ and $U(Y,Z)$; then the composition $V(X,Z)$ is determined by the rule:

$V = T^{\circ}U = max_y\ min\ (\mu_t(x,y), \mu_u(y,z))$

where $\mu_t$ and $\mu_u$ are the membership functions for $T$ and $U$ respectively and $x \in X; y \in Y; z \in Z$.

Composition is used because *NEAR* is a unary fuzzy relation which maps $R \rightarrow R$ and *DISTANCE* is a binary fuzzy relation which maps $R^2 X R^2 \rightarrow R$. Hence:

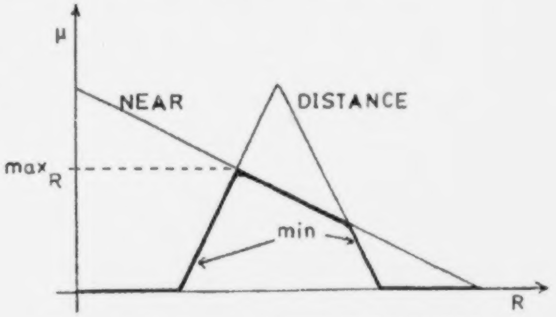


**Figure 6:** *Composition of DISTANCE and NEAR*

$$\mu(DISTANCE \circ NEAR) = max_g \, min\, (DISTANCE(Q,P_f), NEAR(P_f))$$

This is shown in Figure 6. Two observations can be made:

1) as we would expect, the possibility of being *NEAR* increases as centre to centre distance decreases;
2) for a fixed centre to centre distance, the possibility of being *NEAR* also increases as the left spread of *DISTANCE* increases (ie. as the uncertainty in the "shortest" distance increases). We would also expect this to be the case.

## Possibility distributions of spatial relationships

The foregoing has demonstrated how a single exact point $Q$ in $R^2$ can be classified subject to an elastic constraint such as $NEAR(P_f)$. If we now classify each point in a gridded plane subject to this constraint, we can identify areas which have a higher possibility of being *NEAR* to $P_f$. This is a region surrounding the fuzzy point as Figures 7 and 8 show.

Where more than one fuzzy point is involved we can determine if $Q$ is *NEAR* to *all* of them by evaluating the conjunction of the compositions. Thus, with two fuzzy points: $P_{1f}$ and $P_{2f}$, we are evaluating the proposition:

$$S \rightarrow Q \text{ is } NEAR(P_{1f}) \text{ AND } NEAR(P_{2f})$$

for all $Q$ in the grid. The conjunction is evaluated by simply taking the ***minimum*** value of each composition. Figures 9 and 10 show a possibility distribution for points which are simultaneously *NEAR* to two fuzzy points.

We can extend the utility of fuzzy predicates by adding *hedges* (Negoita 1985) such as *NOT, VERY, QUITE*. These modify the shape of a TFN in a manner consistant with their linguistic interprepation. Hence we can create propositions such as:

$$S \rightarrow Q \text{ is } NEAR(P_{1f}) \text{ AND VERY } NEAR(P_{2f})$$

Figure 11 shows a possibility distribution for this based on $P_{1f}$ and $P_{2f}$ in Figure 9.

As an example of how we might apply this; suppose we wished to assess the suitability of an area for the development of a rubber plantation in a developing country. We know that the land should possess a combination of certain physiological and socio-economic characteristics such as soils, present land use and availability. We also require the region chosen to possess a spatial relationship with other features in the area. For example, it should be *near* to a settlement, *not too far* from a port and there should be a road *very near*. We would compute the possibility distribution subject to these spatial constraints and view this as a new attribute to be combined with the others in the analysis.

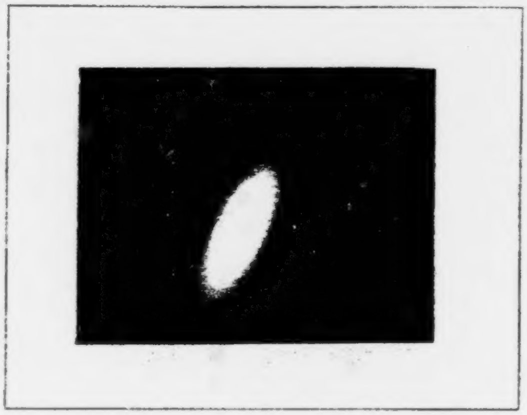


**Figure 7:** *The fuzzy point* $P_f$

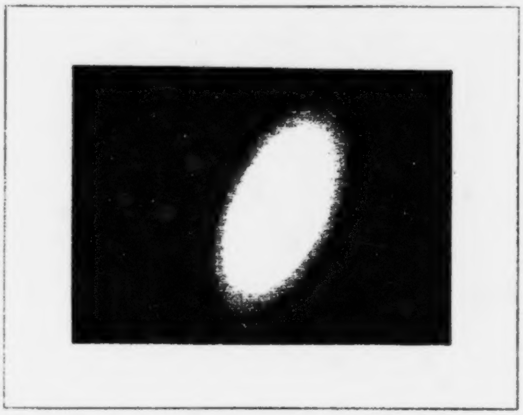


**Figure 8:** *Possibility distribution of* $NEAR(P_f)$

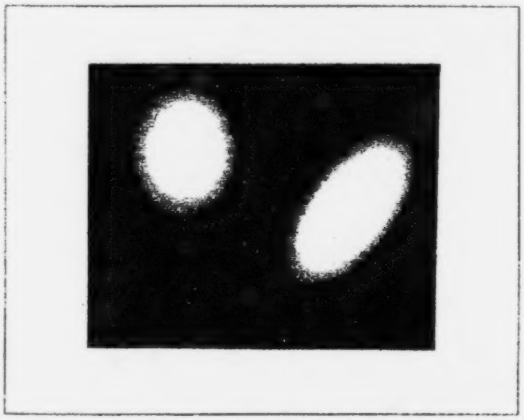


**Figure 9:** *The fuzzy points* $P_{1f}$ *and* $P_{2f}$

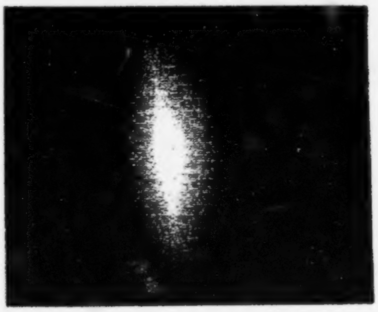


**Figure 10:**
*Possibility distribution of* $NEAR(P_{1f})$ *AND* $NEAR(P_{2f})$

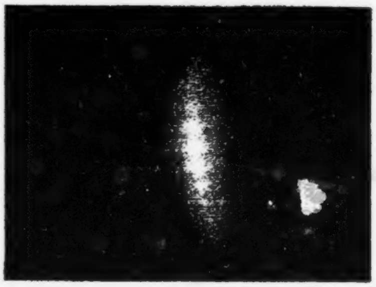


**Figure 11:** *Possibility distribution of* $NEAR(P_{1f})$ *AND VERY* $NEAR(P_{2f})$

**Generalised inexact Spatial Reasoning**

In the general case we have a database of *m* fuzzy points:

$$\varnothing = \{P_1, P_2, P_3, \ldots P_m\}$$

(for clarity we omit the *f* subscript) and a database of *n* unary fuzzy predicates:

$$\Omega = \{R_1, R_2, R_3, \ldots R_n\}$$

The predicates act on some of the points to form *p* constraints of the form:

$$C_k = R_i(P_j) \qquad 1 \leq k \leq p \; ; \; 1 \leq j \leq m \; ; \; 1 \leq i \leq n$$

These constraints can be conjunctively linked into propositions of the form:

$$P \text{->} Q \text{ is } C_1 \text{ AND } C_2 \text{ AND} \ldots \text{AND } C_p$$

and the truth value of the proposition is given by:

$$\mu(Q) = \min_p (\max_k \min(D_k, C_k))$$

where $D_k = DISTANCE(Q, P_j)$ and $P_j$ is the fuzzy point predicated by constraint $C_k$.

Since *Q* can be any point in the plane a possibility distribution for an area can be created. This distribution is a classification of points in the area according to their conformity to the proposition.

## Conclusion

This paper has presented an approach for handling uncertainty in spatial database applications. The uncertainty under consideration originates from errors or vagueness in spatial delineation or from intentional vagueness as expressed by a user in spatial analysis.

We have shown how points in $R^2$ can be classified subject to a set of fuzzy spatial constraints which are joined by conjunction. The resultant classification is a two-dimensional possibility distribution in which the higher values indicate greater conformity to the proposition expressed by the constraints. This classification can be considered a derived layer in a database and combined with other data.

Future work will aim to extend these concepts in a number of ways: defining inexact spatial primitives for lines and regions; extention of the range of spatial relationships which can be accomodated; inclusion of fuzzy attributes; and accomodation of the disjunction operator in propositions.

## References

Dutta, S.(1991). *Topological constraints: a representational framework for approximate spatial and temporal reasoning.* Lecture Notes in Computer Science No. 525, Gunther & Schek (Eds), Springer-Verlag.

Egenhofer, M.J.(1991). *Reasoning about binary topological relations.* Advances in Spatial Databases, Springer Verlag Lecture Notes in Computer Science,525.

Graham, I. & Jones, P.L.(1988). *Expert systems - knowledge, uncertainty and decision.* Chapman and Hall.

Halter, A.N. and Dean, G.W.(1971). *Decisions under uncertainty with research applications.* South-Western Publishing Co.

James, M.(1985). *Classification algorithms.* Williams Collins.

Kaufmann, A. and Gupta, M.M.(1988). *Fuzzy mathematical models in engineering and management science.* North-Holland.

NCGIA - National Center for Geographic Information and Analysis(1989). *The research plan of the National Center for Geographic Information and Analysis.* International Journal of Geographical Information Systems,3(2) pp117-136.

Negoita,C.V.(1985). *Expert systems and fuzzy systems.* Benjamin Cummings.

Peuquet, D.J. (1986). *The use of spatial relationships to aid spatial database retrieval.* Proceedings of the Second International Symposium on Spatial Data Handling, pp 459-471.

Van Roessel, J.W. (1987). *Design of a spatial data structure using the relational normal form.* International Journal of Geographical Information Systems,1(1) pp33.

Zadeh, L.A.(1988). *Fuzzy logic.* IEEE Computer, 21(4) pp83-93.

# D4 Evolving methodologies

**Resource mapping and traversing with the Global Positioning System in northern and central British Columbia**

Jamie McLennan et al., Northern Forestry Mangement, Prince George, BC, Canada

**Creating accurate GIS databases with digital photogrammetry**

Stephen Sperry et al., ERDAS, Inc., Atlanta, GA, USA

**An advanced software engineering environment for prototyping and development of GIS applications**

John Smymew, GeoVision Corporation, Ottawa, ON, Canada

**Analysis of GIS user requirements within the Yukon Department of Renewable Resources using CASE tools**

William Johnstone et al., Crippen Consultants, Vancouver, BC, Canada

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Resource mapping and traversing with the global positioning system in northern and central British Columbia

**Jamie McLennan**
President

**Brian Reeves**
Systems Manager
Northern Forest Management
1967 Ogilvie Street
Prince George, BC V1H 1A2

## Abstract

*The Global Positioning System (GPS) uses a network of NAVSTAR ranging satellites. GPS receivers can generate accurate positions enabling economical traversing and mapping of the forest resource. The results of numerous GPS traverses, completed under a variety of conditions are presented. The results indicate acceptable area and closure errors, and reduced interpretation errors, found in conventional aerial photography mapping methods. GPS data and maps can be presented to GIS users in acceptable formats. GPS is an excellent measuring and mapping tool for resource managers with good potential for other applications.*

## Introduction

The Global Positioning System, or GPS, uses a network of 16 NAVSTAR military satellites. These satellites broadcast very accurate signals and reception from four such satellites can result in an accurate position in latitude, longitude and elevation.

After July 1 of 1991, the U.S. military began a program of inducing errors, called selective availability or SA, into the timed signal, producing up to 100 meters of error in positioning. The effect of SA, can be removed with a base station; a second receiver at a known position which records these military "lies". This data can be applied against field data to produce accurate positions in a process known as differential correction.

Studies of GPS for resource use, conducted by researchers at the Lubrecht Experimental Forest (School of Forestry, University of Montana) demonstrated that a GPS receiver could be used to traverse areas to within an average 2 meter horizontal accuracy in the open and an average five meter horizontal accuracy under the canopy[1]. The U.S. Forest Service has published traverse results[2], using similar GPS equipment, and obtained area errors (when compared to traditional methods of 0.28%). These and other studies inspired recent sales of over 350 receivers to the U.S. Department of Agriculture. The same study also described the use of helicopters to obtain GPS data.

Northern Forest Management uses Trimble Pathfinder Community Base Stations (located in Prince George and Kelowna) and Trimble Pathfinder Professionals. The base stations provide corrective coverage to most of British Columbia, and parts of Alberta, Idaho and Washington State. The manufacturer claims that 2-5 meter accuracy is possible under certain conditions.[3]

## Methodology

### Ground traverse

Nine cutblocks in the Ootsa Lake, B.C. area required traversing before cutting permits could be issued by the B.C. Ministry of Forests. The forest is a Pine-Spruce canopy of 60 to 80% canopy closure.

The receiver was taken to a logical starting point on the cutblock and a data file opened. The receiver was programmed to collect one position per second under specific conditions: three dimensional points only, PDOP ≤ 8.0 and angle mask ≥ 10.0. The operator walked the ribboned boundary and established a numbered station every 200 meters by blazing a tree and noting the approximate position on a field map. The cutblock was walked and the receiver turned off when the start point was reached. The data was off loaded to a laptop PC. Later, this data was deferentially corrected with Base Station data to eliminate SA and other errors. A closure error was calculated by measuring the distance from the starting point to the end point. This is not truly a closure error in the traditional sense, when errors have accumulated over the whole of a traverse. GPS positions are generated independent of previous positions, so the calculated difference is really a measure of repeatability and an indicator of the quality of our results. Table 1 describes the traverse length and "closure error" in meters and percent for each block traversed.

The simple average of all closure errors was 7.48 meters, ranging from 2.42 meters to 12.77 meters. The simple average closure was 0.146% (ranging from 0.50% to .308%). The operator of the system reported that he could complete approximately 1 block or 5 km each day, slightly more than a two person crew could accomplish in the same time. Actual costs were similar to conventional traversing. Figure 1 illustrates the positions found on typical block.

**Table 1**
**Under canopy GPS traverses, October 1991**
**Forest licence a16826 Ootsa Lake, B.C..**

| Block | Gross Area (ha) | Net Area (ha) | Perimeter Block (m) | "Closure" Distance (m) | "Closure" Error (%) |
|---|---|---|---|---|---|
| 302 | 80.9 | 80.9 | 4806.41 | 7.92 | 0.165 |
| 319 | 84.1 | 90.9 | 6522.9 | 9.26 | 0.142 |
| 311 | 49.8 | 49.8 | 3657.5 | 2.42 | 0.066 |
| 307 | 135.9 | 120.5 | 4935.1 | 12.91 | 0.149 |
| 30 | 128.1 | 128.1 | 6291.8 | 4.82 | 0.077 |
| 326 | 82.3 | 82.3 | 4144.0 | 12.77 | 0.308 |
| 335 | 124.9 | 119.5 | 5608.9 | 3.53 | 0.044 |
| 339 | 27.1 | 25.9 | 2132.8 | 10.01 | 0.315 |
| 329 | 115.6 | 110.9 | 4863.5 | 3.75 | 0.050 |
| | | | **Simple Average** | **7.48** | **0.146** |

**Collection Parameters**

A. Three dimensional data only

B. PDOP ≤ 8.0

C. Differentially corrected from Prince George, B.C.

D. Angle mask ≥ 10.0 degrees

(Source: Northern Forest Management, Prince George, B.C.)

**Figure 1**

*Under Canopy Traversing*

*Block 311*

*Scale 1:20 000*

*Area 49.8 ha*

*Tick 15"*

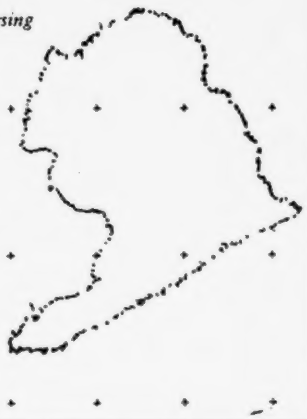


**Figure 2**

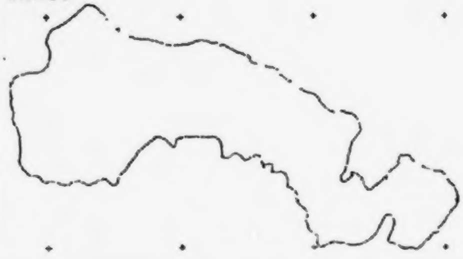
*Aerial Traversing*

*CP 168, Block 100*

*Scale 1:20 000*

*Area 73.0 ha*

*Tick 30"*



## Aerial traversing

Forest licensees are sometimes billed stumpage or other fees based on the area of a cutting permit. The Ministry of Forests checks these areas or requests traverses to be performed that have closure errors of less than 0.4% and area errors of less than 1.0%. In the northern portion of the province, these checks may require helicopter access to remote areas.

A GPS antenna was mounted on a Bell 47 T helicopter. Detailed attention was made to the location of the antenna, so as not to pose a safety problem, nor interfere with reception of satellite signals.

The GPS receiver and recorder were placed inside the helicopter and operated by a navigator. The pilot was instructed to fly the perimeter, keeping the aircraft to under 20km/hr and as low as safely possible to the perimeter trees. The navigator was kept busy noting timber irregularities, such as windfall or beetle attack on the perimeter. He also noted where swamps, creeks and topographical information was and provided guidance to the pilot to exact boundary locations when adjacent to swamps or non productive forest.

Twenty cut blocks were traversed in this fashion for the Ministry of Forests, Mackenzie district. That district also required "error of closure" as a check on repeatability. In addition, they took the initiative to check two of the blocks using a Theodilite survey crew from the Prince George Regional Office. Table 2 records their results and indicates a *very* low area error at 0.12% and 0.13% respectively. Figure 2 shows the typical results from an aerial traverse.

*Safety* is an important concern with this technique. The pilots reported that the technique was not necessarily dangerous, but the avoidance of hazards and prolonged slow and precision flying was tiring. Any wind added to the work load and diminished the safety margin. It was highly recommended, by almost all pilots, that only turbine type helicopters be used for reasons of power, reliability and safety. Proficient pilots and well maintained, powerful helicopters should only be used for this technique.

This technique was very productive, and up to 60 kms per day, or 10 cutblocks, could be traversed in this manner. The productivity and cost effectiveness were very sensitive to travel or "ferry time". If this was minimized, a cost savings of 20% from conventional methods may be possible.

## GPS vs aerial photography

The Ministry of Forests, Prince George District compared ground and aerial GPS traverses against maps produced from aerial photographs. The GPS traverses were completed as previously discussed and the aerial photography interpretation used a Zeiss Topocart stereo plotter. The aerial photos were used, in stereo, to plot boundaries, roads, landings and adjacent features.

The overall results produced a variety of differences, generally based on interpretation differences (see Table 3). In general, aerial photography will tend to include small, and sometimes large, portions of non-productive (NP) forest. These portions of NP forest may lie on the boundary and are mistakenly included as being within the logged area. NP may also be misinterpreted, or excluded within the block. The GPS operator, usually a forest technician, has the advantage of being able to make on site decisions as to the location of the boundary. The photogrammetrist must make a decision as best as his experience and the photographs allow. This situation is best illustrated by TSL A27894 where a difference of 2.6 hectares is indicated. The GPS operator noted an old slide against the boundary and correctly excluded it from the traverse. The photogrammetrist could not distinguish the slide from the boundary, and included it in the

## Table 2
## GPS Aerial traversing
## Ministry of Forests
## Mackenzie District
## May 1991

| CP | CB | Measured Area (ha) | Measured Perimeter(m) | Error of Closure (%) | Area Error (%) |
|---|---|---|---|---|---|
| 165 | 178 | 20.1 | 1,952 | 7.0m (0.36%) | |
| 167 | 219 | 153.2 | 7,328 | 10.5m (0.14%) | 0.13% |
| 168 | 100 | 73.0 | 5,616 | 7.4m (0.13%) | |
| 211 | 451 | 160.2 | 8,439 | 14.8m (0.17%) | |
| 211 | 456 | 71.0 | 3,812 | 14.8m (0.38%) | 0.12% |
| 253 | 721 | 131.8 | 6,192 | 7.4m (0.11%) | |
| 307 | 286 | 100.4 | 7,073 | 7.4m (0.10%) | |
| 337 | 291 | 77.5 | 4,392 | 10.0m (0.23%) | |
| 339 | 1346 | 100.2 | 6,002 | 14.8m (0.24%) | |
| 367 | 5583 | 27.2 | 2,999 | 9.6m (0.32%) | |
| 371 | 5510 | 120.4 | 6,104 | 7.0m (0.11%) | |
| 463 | 109 | 67.2 | 6,720 | 14.8m (0.21%) | |
| 467 | 154 | 177.3 | 10,933 | 1.4m (0.01%) | |
| 508 | 149 | 37.0 | 3,850 | 7.4m (0.19%) | |
| 509 | 1 | 154.2 | 5,536 | 10.5m (0.19%) | |
| 562 | 311 | 155.6 | 7,397 | 16.5m (0.22%) | |
| 563 | 312 | 80.0 | 4,763 | 0.5m (0.21%) | |
| 170 | 200 | 56.8 | 6,981 | 7.4m (0.10%) | |
| 214 | 3454 | 111.4 | 5,987 | 14.8m (0.24%) | |
| 260 | 4809 | 104.0 | 4,388 | 13.0m (0.29%) | |
| TOTAL | | 1,978.5 | 116,464 | | |
| **Simple Average** | | | | 9.85m (0.19%) | 0.125% |

**Data collection parameters**

A. 1 position per second

B. PDOP $\leq 8$

C. Differentially corrected from Prince George

D. Angle mask $\geq 10$ degrees

logged area. The difference between the two methods is only 0.4 hectares when the slide area is eliminated from the photo.

All the GPS traverses were smaller in area than the aerial photography interpretations. It is possible that shadows, slope distortion or parrallax played a part in the differences.

Ministry of Forests personnel were "inclined to accept GPS as more accurate" than maps interpreted from aerial photographs.

## The effects of selective ability (sa)

Selective Availability (SA) became clearly noticeable after November 15, 1991. Autonomous (uncorrected) GPS data now contains up to 100 meters (and more) of military "lies". Figure 3 shows the effects of SA on cut blocks traversed on the ground. Note the "circle" of points in the northwestern portion, and the jump of positions in other areas. Unless our ground personnel were into the "wildwood weed" (and I am assured they are not!), we are certain to be witnessing extreme SA error. Figure 4 shows the corrected results having used base station data to differentially correct the field data. The uncorrected data cannot be used to even approximate the actual block size or shape. GPS accuracies of 2-5 meters can only be obtained using differential corrections.

## GIS

### Database

GPS is an excellent tool for loading and updating GIS spacial databases. In the very near future, we will be able to tag spacial data with attributes in the field to provide increased functionality with minimal manual intervention. A turnkey GPS to GIS system will be a reality in 1992.

**Table 3**

**GPS Traverse areas vs stereo aerial photography** 4

| Block GPS | Area (ha) | Aerial Photography | Difference (ha) | Ministry of Forests Comments |
|---|---|---|---|---|
| TSL A27921(a) | 71.9 | 73.9 | 2.0 | Possible interpretation errors requires ground check. |
| TSL A27894(a) | 53.2 | 55.8 | 0.4 | A/P included 2.2 ha of old slide, not part of block. Actual A/P area approx. 53.6 ha. |
| TSL A27973(g) | 41.3 | 43.6 | 2.3 | Interpretive errors, GPS seems more accurate upon closer examination of photo. Shadows on photograph make A/P interpretation difficult. |
| TSL A27973(g) | 12.1 | 12.9 | 0.8 | Interpretive errors inclined to accept GPS. |

**GPS traverse method: (g)=ground, (a)=aerial**

**Overall Comments**

- more detail from aerial photograph
- inclined to accept GPS as more accurate
- field checking would clarify "interpretive errors"
- GPS avoids significant "miss" (ie. NP, slides or blowdown)

## The effects of selective availability

**Figure 3**

*Atonomous Data*

*-1 postion/2 sec.*

*-3D*

*-Elev. Mask 8 degrees*

*-PDOP ≤ 8*

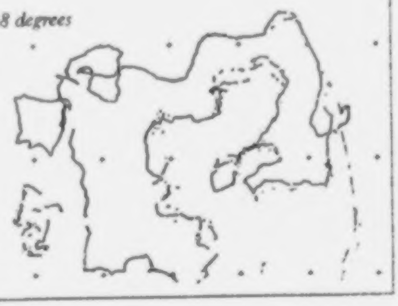


**Figure 4**

*Corrected Data*

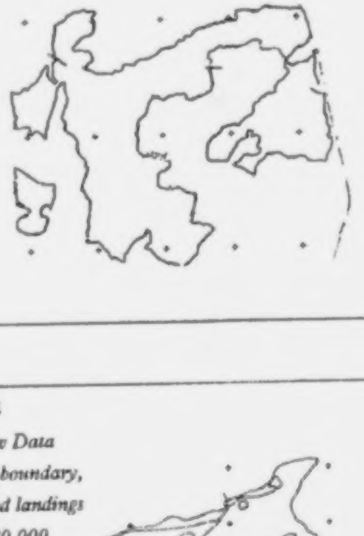


**Figure 5**

*GPS Raw Data*

*Cutblock boundary,*

*Roads and landings*

*Scale 1:20 000*

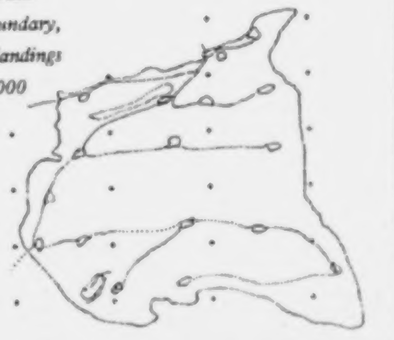


**Figure 6**

*GPS Raw Data*

*Converted to GIS*

*Scale 1:20 000*

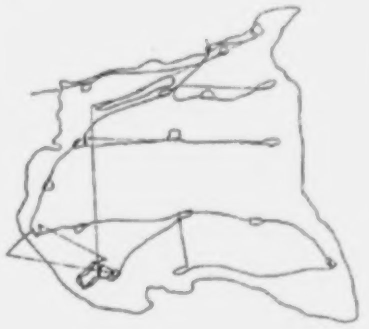
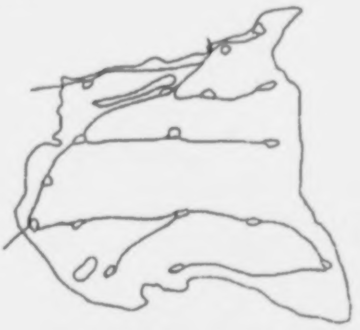


**Figure 7**

*Layered GPS*

*Data converted to GIS*

*Scale 1:20 000*



### Data collection

GPS data collection techniques may have to be modified for GIS applications. Since each point collected in GPS is independent of every other point, the system allows you to collect data in any order you wish (Figure 5). For example, on day 1, you could digitize one side of a cutblock, then do some roads and landings. On day 2, finish the block and then finish the roads and landings. GIS would have problems making sense from this data (Figure 6). A simple solution would be to store the cutblock boundary information in one file and the roads and landings in another file (Figure 7).

## GIS formats

GPS data can be converted to the following GIS formats:

- ERDAS
- Autocad DXF
- Arc-Generate
- Grass Sites
- Grass Vectors
- Gridpoint EPPL7
- P.C. Moss
- McDonnel Douglas GDS
- USGS DLG
- SEG-P1
- IBM GFIS
- Others

We are generating maps exclusively in NAD-83, although NAD-27, WGS-84 and other datums can be used. Coordinates can be generated in 2D or 3D, (although we always collect 3D data), in Latitude/Longitude, Universal Transverse Mercator or State Plane coordinate systems. Three D data may be generated with Z's given as height above ellipsoid (HAE) or mean sea level (MSL). GPS is based on the WGS-84 Ellipsoid and conversions of Z values to HAE are approximate and can introduce a few meters of vertical error.

## GIS conversions

Conversions from GPS to any of the above GIS formats are straight forward. However, our experience with converting maps from one GIS format to another led us straight to the local drug store to buy more aspirins.

Conversions are never as straight forward as software manufacturers would have you believe.

In the past, we were creating DXF (Drawing Interchange File) files in Autocad for a client that was using Microstation. After many failed attempts to load our conversion we found it is impractical to convert from one piece of software to the other. In fact, we failed to load those same DXF files created in Autocad back into Autocad. Our aquisition of a GIS system will make conversion easier.

## Discussion

The results of the paper parallel those found in other GPS studies and demonstrate the usefulness of GPS to forestry applications. In under the canopy work, GPS can produce acceptable results; the observed closing error (simple average) of 7.45 meters is slightly higher than the manufacturer's claim of 5 meters. The difference is most likely accountable to multipath, or signal reflections, which introduce a small additional error. At the time of the surveys, the canopy and tree boles were reported to be wet and this may increase the number of reflections. It must be noted that not all canopies may permit reception. GPS signals are "line of sight" and cannot be received under certain conditions in the canopy.

Traversing with GPS under the canopy, and in the open can be more accurate and productive than conventional traverse methods.

Helicopter traversing is an emerging technique with great potential. Lines may be traversed accurately at an unprecedented rate. The onboard navigator is able to minimize pilot error and provide assistance in determining actual boundary and other features. The results certainly indicate acceptable closure and area errors.

The helicopter method may also include some hover error, which by observation may be in the neighbourhood of 1-2 meters depending on wind conditions at the time. This kind of traversing for cutblocks, fires, road systems, site treatments, insects and disease will probably offer savings and increased accuracy in the future.

GPS may offer increased accuracy over maps from aerial photography by the reduction of interpretation errors inherent in aerial photographs. Although the GPS results correspond closely with the photograph results, GPS may eliminate interpretaion errors undetectable from aerial photographs.

Selective availability, the introduction of military lies, produces errors in the 100 meter range. These errors must be corrected by use of a base station, known as differential correction, if acceptable accuracies are to be obtained.

Data conversions from the GPS Pathfinder Professional Software to GIS formats is straight forward. Data collection methods will have an impact on the quality of the data presented to the GIS system.

## Acknowledgments

[1] Gerlach, Fredrick, *Global Positioning System - Input to GIS.* Society of American Foresters, Spokane Washington, September 27, 1989.

[2] Bergstrom, Gary; *GPS in Forestry Management*, GPS World Magazine, October 1990.

[3] Trimble Navigation Ltd. et al, *Geographic Information Systems and the GPS Pathfinder System.* December 1990.

[4] Source: Ministry of Forests, Prince George District, Small Business Forester, Summarized Data

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Creating accurate GIS data bases with digital photogrammetry

**Kurt Novak**

Ohio State University
Department of Geodetic Science and Surveying
1958 Neil Avenue
Columbus, Ohio 43210-1247 USA
Tel: 614/292-6753

**Stephen L. Sperry**

ERDAS, Inc.
2801 Buford Highway, Suite 300
Atlanta, Georgia 30329-2137 USA
Tel: 404/248-9000
Fax: 404/248-9400

### Abstract

*This paper investigates the full integration of digital photogrammetric functions in a geographic information system (GIS). Photogrammetry is a very powerful tool that can be used to create accurate data for a GIS. The major obstacle against a wider application as a data capture method is the complicated instrumentation necessary. However, an implementation of photogrammetric mapping and triangulation functions on a workstation in an image analysis environment eliminates the cost of analytical instruments. This makes the system easy to use for non-photogrammetrists and it can be done directly in a data base environment.*

*The concept of a digital photogrammetric plotter implemented as a software module of an existing GIS is presented. A full range of photogrammetric functions was developed in a workstation environment. They can be used to generate digital elevation models (DEMs) and digital orthophotos from aerial photographs and satellite imagery. These are two important layers of a GIS. The developed software is not limited to digital mapping, but can be used for highly accurate aerial or satellite triangulation. Stereo images are displayed on a workstation screen where points can be selected semi-automatically. This allows analysts to accurately integrate satellite data with digitized aerial photographs in a common system.*

## Introduction

Although the geographic information system (GIS) has become a popular tool for many organizations, their implementation is often stalled by the lack of available digital data. Most GISs are based on vectors (line data) digitized from existing maps. The digitized maps often include error-those transposed from incorrect or out-of-date maps, as well as those occurring during digitizing. Two critical GIS data layers, terrain and land cover, are typically difficult to develop in vector format. Contour maps representing terrain leave many areas blank. Land cover maps generated through manual air photo interpretation methods are time consuming and expensive. These data gaps can be filled by information from a raster GIS.

Information for a raster GIS is mostly derived from satellite sensors or scanned aerial photographs which record spectral information about the surface of the earth. The raster equivalents of vector contours and land cover maps are digital elevation models (DEMs) and digital orthophotos. Digital orthophotos are image-based maps which have been corrected for relief displacement.

In the past, DEMs and orthophotos were derived by manual photogrammetric techniques. Through the development and enhancement of digital photogrammetry, it is possible to derive the information fully automatically from digital imagery. Because of the similarity in data formats between digital photogrammetry and GIS, the DEM and digital orthophoto can be readily used in a raster GIS. Due to recent advances in the integration of raster and vector GIS, the data are also available to vector a GIS.

Digital photogrammetry provides not only current and accurate information for a GIS, but its algorithms are also useful for extracting products from existing GIS data layers. These products include perspective views, slope models, exposition models and contours. This paper describes the development of a digital photogrammetric module in a raster GIS. First, the functions by which digital photogrammetry can enhance the capabilities of automatic information extraction are discussed. This is followed by a discussion of a specific implementation of photogrammetric algorithms in a GIS. The discussion includes descriptions of sensor orientation, image matching, DEM interpolation and digital orthophoto generation. Typical problems that are encountered in these processes are identified along with operator solutions. In the conclusions, future developments and further enhancements of digital photogrammetry are discussed in terms of new photogrammetric image analysis techniques.

## Digital photogrammetry

Photogrammetry has been used for about one hundred years to generate maps from aerial photographs and to compute highly accurate point positions in three dimensions. It is a matured technique for extracting spatial data. The advance of digital sensors have made it possible to use photogrammetry on a computer and apply its algorithms to digital imagery. The use of a computer in photogrammetry is known as a softcopy photogrammetry system. Currently, the most popular techniques are image matching and DEM generation, as well as digital orthophotography.

Image matching can be classified into different techniques depending upon the features used to detect similarities. The most popular is area-based matching in which a gray value matrix of one image is compared to a gray value window of another image on a pixel by pixel basis. This method is most accurate for measuring image coordinates of well defined points as it gives sub-pixel accuracy. However, this method also requires very good approximations (about 5 pixels) in order to find matches successfully. Other matching techniques are based on features which could be lines or characteristic points of the image. They must first be extracted before the matching can begin. This technique is less accurate, but more robust since correct matches can be found over the whole image area without any approximations.

Other types of image matching, such as binary tree matching, are currently under development and have not been practically implemented in digital photogrammetric systems. The matching of points can take place in image space (on the sensor) or in object space (on the map or on the ground). The relationship between image and object is given by the collinearity (perspective) equations, which form the basis of all photogrammetric point positioning methods. Any point-pair matched in image space can be immediately projected to the ground.

Once in object space, the 3-dimensional points can be used to generate a raster DEM stored in the same format as digital image (a matrix of gray values). Photogrammetry has been involved in the development of DEM interpolation techniques for a long time. Good DEM interpolation techniques allow us to approximate the terrain smooth functions by using reference points. It is very important to smooth reference points derived automatically in order to eliminate wrong matches and reduce noise. Analytical surfaces developed during the DEM interpolation process are represented as a dense grid of elevation points in the GIS. At each location of an elevation pixel, the corresponding gray-value can be found by digital orthophotography. The surface point represented by the DEM pixel is projected into the original image using the perspective relationship between image and ground. The location in the original image is identified and its corresponding gray value is extracted by a resampling technique. Once completed over the whole image, the digital orthophoto is created. This layer fully corresponds to a (digital) map free of any relief displacements.

Digital photogrammetry is heavily involved in the development of feature extraction and image understanding techniques. The digital mapping of linear features such as roads, rivers or rectangular objects such as buildings is of profound importance. Research is currently underway to automatically find these lines, which are usually highly visible in images and supposedly correspond to edges of the real surface. Once lines have been detected and vectorized, artificial intelligence has to be applied to automatically interpret their meaning. Image understanding is a popular discipline which will soon allow us to fully and automatically analyze vector data obtained from the feature extraction. This data can then be integrated into the GIS together with the raster information.

## Implementing digital photogrammetry into a raster GIS

Many of the techniques described in the previous section were implemented in an existing raster GIS to allow the user to acquire spatial data directly in the familiar GIS environment using digitized aerial photographs or satellite imagery. They are collected in a digital photogrammetric module, mainly applied for information extraction for the GIS data base. The functions are comprised of coordinate measurement in the imagery, control data collection, sensor orientation for aerial photos and satellite imagery, image matching, DEM interpolation and orthophotography. The user can densify control point networks by aerial triangulation or satellite triangulation, create a digital elevation model by image matching and DEM interpolation, and consequently derive a digital orthophoto by using the DEM to correct for relief displacements. Figure 1 below shows a flow chart describing all of the functions implemented in this module.

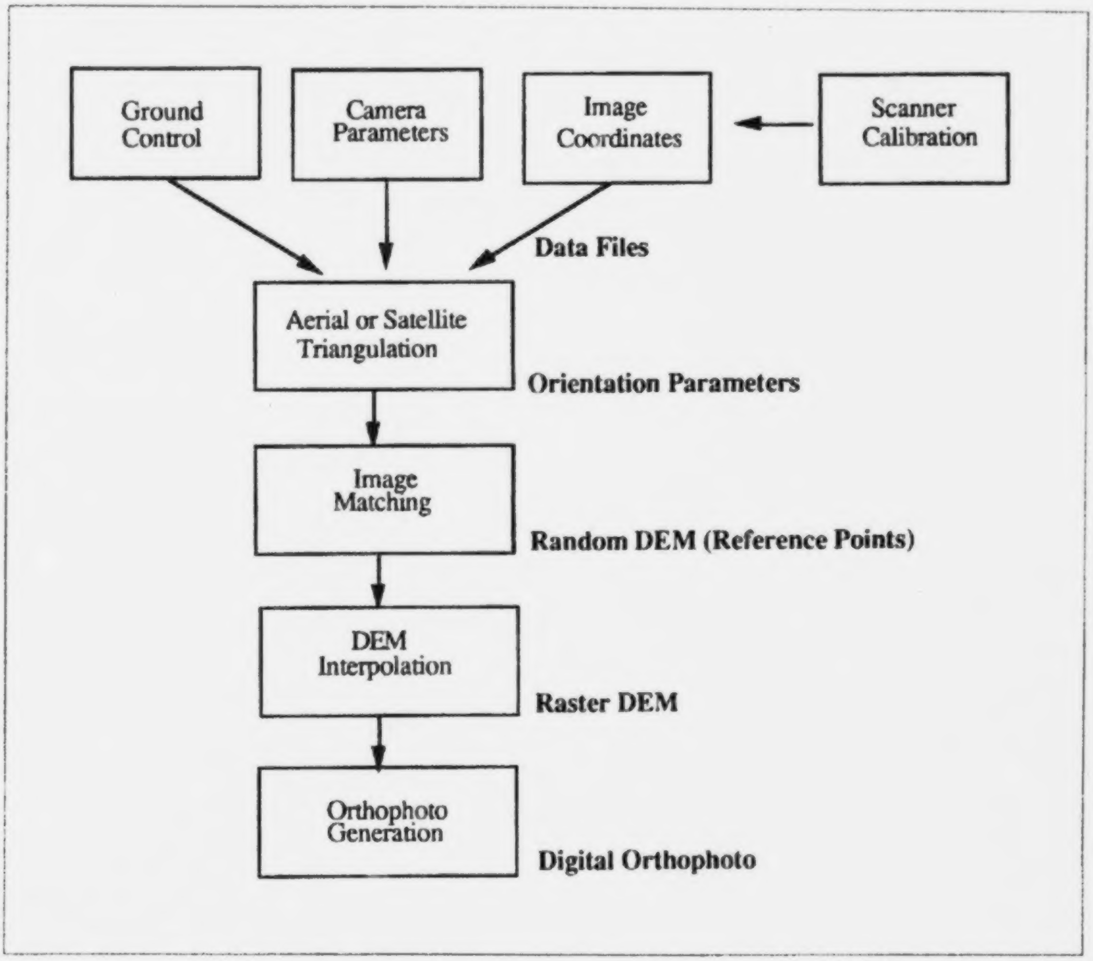

Ground Control
Camera Parameters
Image Coordinates
Scanner Calibration
Data Files
Aerial or Satellite Triangulation
Orientation Parameters
Image Matching
Random DEM (Reference Points)
DEM Interpolation
Raster DEM
Orthophoto Generation
Digital Orthophoto

Figure 1: This diagram shows the digital photogrammetric functions integrated in a raster GIS. The boxes indicate the functions and the text to the right state to the data derived at each stage.

In the following sections, the typical application of the photogrammetric module is described.

### Image coordinate measurement

The first step of data analysis is the preparation of the data files needed for the sensor orientation and as approximations for image matching. Among these data files are a list of ground control points, image coordinates, camera calibration information, and, if available, approximations of the sensor positions and attitudes.

Two windows are used to display a digital stereo image pair on the computer screen in a reduced form. This is necessary as the digital images very large (6000 x 6000 pixels for SPOT imagery and 4000 x 4000 pixels for digitized photographs). A small window can be enlarged to full resolution so that the operator can identify and measure image coordinates to within a pixel, or by using a resampling technique, even to a quarter of a pixel. These measurements are done manually on the screen. At the same time, the operator identifies control points on an existing map. This map must be accurate relative to the digital imagery (e.g., USGS quad sheets at 1:24,000 can be used for SPOT imagery). Otherwise, ground control points must be established by geodetic or terrestrial survey techniques. The split screen simulates the functionality of a stereo comparator to select points in the left and right images of the stereopair. It also allows the user to form a strip of photographs by shifting from one stereopair to the next on the monitor. If aerial photographs are used, the pixel coordinate measurements are transformed into an image coordinate system. This coordinate system is defined in the calibration report by the camera manufacturer, which is available to the user of the aerial photographs. The split screen tool can later be used to edit match points in the images.

The user has an option to check the accuracy of the scanner used for digitizing the aerial photographs. The scanner calibration program was developed to model the distortions of the scanner and to apply corrections to the image coordinates. This is an important procedure because many scanners apply a second projection to digitize the images, in which case additional distortions are unavoidable. These distortions are often larger than 3 pixels.

### Sensor orientation

The sensor orientation is typically done by photogrammetric bundle triangulation. This technique is based on collinearity equations and is widely used in analytical photogrammetry for aerial triangulation. It is a very flexible tool used to solve for the exterior orientation parameters of a photograph (the exposure station and attitude of the camera), to densify the ground points, to derive camera parameters, and to model distortions of the sensor system. The bundle triangulation technique can not only be used for aerial photographs which are modeled by a central perspective geometry, but it can also be modified for satellite sensors. Satellites use separate scan lines that are acquired sequentially. Each of these scan lines has a separate perspective center. Due to its smooth motion in orbit, all perspective centers can be lined up along an analytical function. Therefore, the geometry of the satellite sensors is fundamentally different from aerial photographs. However, it can also be modeled by a modified bundle solution.

Both of these functions were implemented so that the user can compute the orientation parameters of either aerial photographs or satellite scenes. Furthermore, the user can densify points on the ground which were not digitized in the maps network densification. As they are positioned very accurately, they can be used as a reference in separate images. Sensor orientation parameters are important for 3-dimensional positioning of points to create the random DEM, and for digital orthophotography.

### Image matching

The image matching techniques implemented are following the area-based approach. To minimize the need for approximations, an image pyramid is created by reducing the original image four times by a factor of two. At the lowest resolution, a square grid is established in the left image and matched to the right one. Once this grid has been found, its points are projected to the next higher level of the pyramid (higher resolution) until the match points finally appear in the full resolution image. This first, coarse grid can be edited on the split screen monitor so that the user can correct wrong matches. These wrong matches typically occur in areas of low contrast such as lakes or desert areas. Afterwards, a densification is performed to obtain a dense set of reference points in the images. For example, in SPOT scenes points can be matched as close as every third pixel.

Two area-based matching techniques are applied: cross correlation and least squares matching. Cross correlation finds corresponding image points at an accuracy of one pixel. It is a very robust technique which locates the points without very precise approximations. For the densification, least squares matching is used. This technique yields an accuracy of a tenth of a pixel in the image.

Once these sets of corresponding points are available and have been edited, an intersection is computed to project the image coordinates into object space by virtue of the known orientation parameters of the two images. In object space they form a random DEM of reference points which can be used to interpolate a grid DEM, or they can be connected by triangles using a TIN structure.

### DEM interpolation

This procedure transforms the spatial random points to a regular raster format. It can also filter elevations to eliminate wrong matches. The algorithm implemented is based on the summation of surfaces, which creates a smooth, analytical surface over the whole interpolation area. Wrong matches stick out of the surface as spikes. These areas and other noise are filtered at this stage to fit the surface to the reference points in the best way. The surface of this algorithm also covers areas with or without sparse reference points. As a result, a raster DEM is generated that directly forms the elevation layer of the GIS. It is georeferenced and can be created in any selectable map projection supported by the GIS.

### Digital Orthophoto Generation

Once a DEM is available, the relief displacement of the original images can be corrected by applying the surface elevations to this image. Basically, any DEM pixel is projected back into the original image where we can resample a gray value. This gray value is placed in the corresponding DEM pixel location in the orthophoto plane. The relief corrected image is georeferenced to the map projection as the DEM. As there is no relief displacement, this image can be directly overlaid with vector data without showing any offsets. This is a typical problem of regular geocoded images that were not relief corrected.

Although the procedure of creating orthophotos and DEMs was described in a sequential form, the digital photogrammetric module is very flexible. The user can enter the flow chart (figure 1) at any stage if the appropriate information is available. The user can also combine data from different sources. For example, if the major objective is to create highly accurate digital orthophotos at a high resolution, they should be produced from large-scale aerial photographs. The DEM used to correct relief displacements need not be very accurate in this case. Actually, it can be derived from satellite imagery, such as SPOT Panchromatic stereopairs. The DEM would still have a vertical accuracy of 10 - 15 meters, which hardly effects the planimetric accuracy of the final orthophotos. Thus, the user can create the raster DEM using SPOT imagery and then apply it to aerial photos for relief correction. The orientation parameters must be derived from both sensors.

## Future directions

### Conclusions

The integration of digital photogrammetry and raster GIS is a very important enhancement of the way in which information is acquired for geographic information systems. Currently, the techniques to automatically derive digital elevation models and digital orthophotos in raster format are operational in the ERDAS GIS package. Many enhancements of this system can be envisioned. For example, it would be useful to use a real, 3-dimensional monitor to view image pairs as stereo-models, in which the operator could trace lines in space with a 3-dimensional cursor. The automatic interpretation of raster data is of critical importance for combining raster and vector GIS capabilities. Therefore, feature extraction algorithms are being developed that will permit the user to automatically find lines in the images, and even interpret them as objects on the ground such as roads, rivers or buildings.

Another enhancement to speed up digital mapping is automatic line following in the digital images. For this procedure, the operator would select the starting point of a linear feature, such as a road or a river, and the function would automatically trace this line as long as it is unambiguous.

In the future there will be many new developments in automatic data acquisition for geographic information systems. Software vendors have recognized the importance of integrating photogrammetric functions in a GIS and are working on modules similar to the one described here. Ultimately, this information extraction technology will also enhance the importance of satellite sensors and digital aerial cameras for small-scale mapping and geographic data bases, as well as for civil engineering applications.

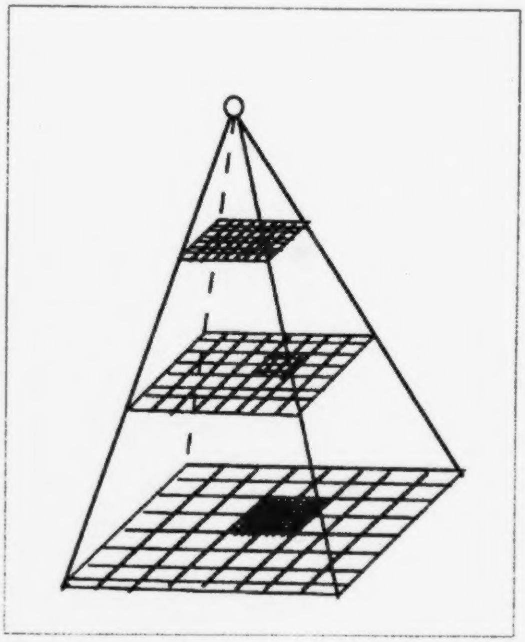


Figure 2

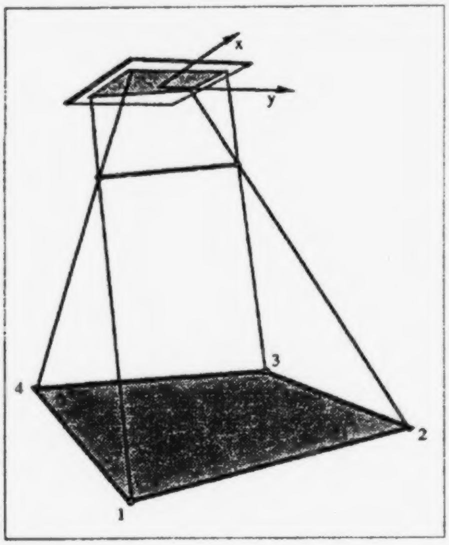


Figure 3

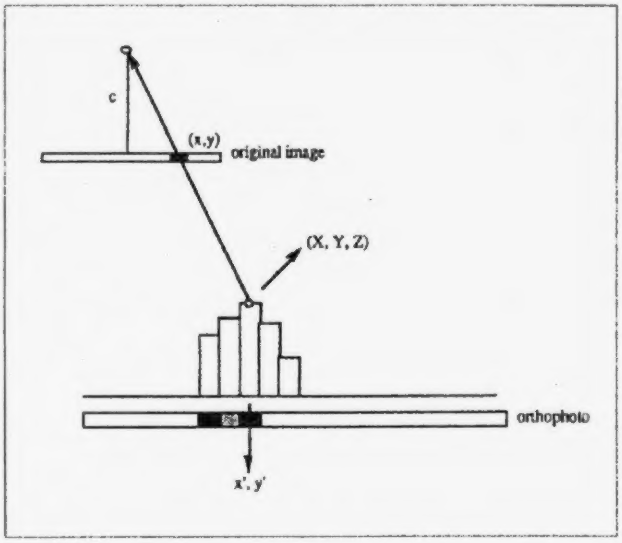

c
(x,y)
original image
(X, Y, Z)
orthophoto
x', y'

Figure 4

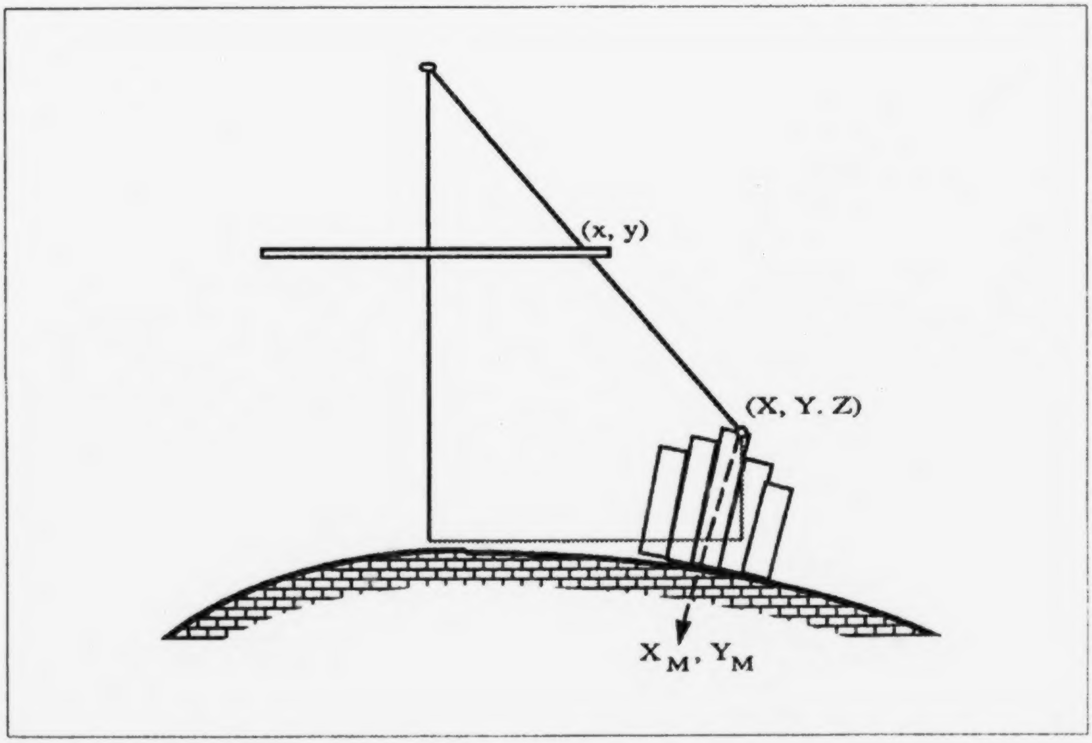


Figure 5

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# An advanced software engineering environment for prototyping and development of GIS applications

**John M. Smyrnew**

Principal Application Specialist
GeoVision Corporation
1600 Carling Avenue, Suite 200
Ottawa, ON K1Z 8R7

## Abstract

*The exploitation of Geographic Information Systems (GIS) technology is heading into its third decade. It is now considered by many an all encompassing base technology similar to Computer Aided Drafting (CAD) and Relational Data Base Management (RDBM) systems used for both traditional and new types of applications. Although there are still requirements for so-called "turn-key" applications, the modern approach is to instead provide an advanced software engineering environment that is embedded into the GIS. This advanced software engineering environment would provide tools at a variety of levels that may be used for prototyping and development of georeferenced and non-georeferenced applications. At the lowest level, a run-time subroutine library accessible by several programming languages and debuggers would be used by computer systems staff. The highest level of tools consisting of psuedo-Computer Aided Software Engineering (CASE) based source code, user-interface, and query generators would be used primarily by end-user and operations staff. In addition to supporting the development of applications, the advanced software engineering environment may also be used for integrating external applications in an enterprise-wide GIS. This paper discusses the new generation of tools required in an advance software engineering environment, their features and limitations, some sample applications, and of course what this means to an organization.*

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Analysis of GIS user requirements within the Yukon department of renewable resources using CASE tools

**William M. Johnstone**

Manager, GIS Applications
Crippen Consultants
A Division of H.A. Simons Ltd.
900-401 West Georgia Street
Vancouver, B.C. V6B 5E1
Canada

**Beth Hawkings**

**Lauren Crooks**

GIS/Remote Sensing Coordinators
Department of Renewable Resources
Government of Yukon
Box 2703
Whitehorse, Yukon Y1A 2C6
Canada

## Abstract

*Crippen Consultants, a division of H.A. Simons Ltd., has recently completed an analysis of user applications and databases for a Resources and Lands Information System (RLIS) for the Yukon Department of Renewable Resources. The RLIS will be a computer-based system comprising two key technologies: a Geographic Information System (GIS) and a Remote Sensing Image Analysis System (IAS). The GIS is seen as a tool to facilitate an evolution towards integrated resource management. It will be used for a wide variety of application areas including: agriculture, fisheries, environmental impact assessment, habitat management, regional planning and land claims implementation.*

*As part of the process of developing a more detailed understanding of the user requirements, Crippen used a Computer-Aided Software Engineering (CASE) tool to assist in the process of documenting user applications and databases. CASE tools are gaining increasing acceptance as an enabling technology that can facilitate the GIS implementation process. This paper will describe the process that was followed in performing the analysis, and discuss the benefits and pitfalls of using this approach to document and analyze user requirements.*

## Introduction

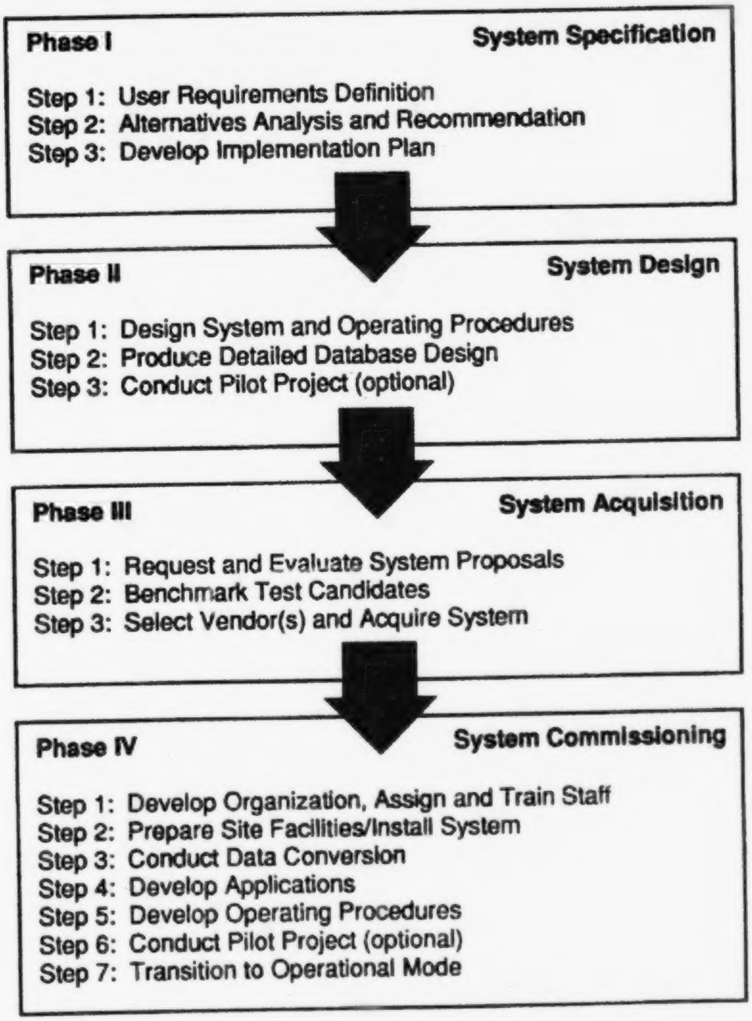
This paper describes the process followed by the Yukon Department of Renewable Resources (DRR) and Crippen Consultants to document user requirements for a Resources and Lands Information System (RLIS) using a Computer-Aided Software Engineering (CASE) tool. It is written for managers and users who are not familiar with CASE tools and who might consider the use of such a tool in their own GIS implementation process.

**1.1 Project history** In March of 1986, Yukon DRR requested that an Information Systems Plan be developed as part of an effort to improve the planning, control, development and management of the Department's programmes. One of the plan's recommendations was to automate DRR's geographically referenced databases. A proposed RLIS would combine Geographic Information System (GIS) and Image Analysis System (IAS) technologies to allow DRR to utilize spatial data in a more efficient manner.

In October of 1989, DRR issued an RFP to develop an executable implementation plan to establish the RLIS. Crippen Consultants, a division of H.A. Simons Ltd., was awarded a contract in December 1989 to develop this plan. The project was divided into four tasks: (1) User Requirements Definition, (2) Alternatives Analysis and Recommendation, (3) Development of an Implementation Plan, and (4) Preparation of a Management Presentation for funding approval. This contract was completed in July, 1990. This process is explained in more detail in Johnstone (1991).

In January 1991, Crippen was awarded a second contract to begin work on Phase II of the implementation process (Figure 1). The purpose of this new contract was to begin documenting user requirements in detail. Once Phase II is completed, DRR will have a System Design Document which describes the RLIS in terms of: system architecture, user applications, global data architecture, detailed functional requirements, and operational issues. This document will form the basis for the acquisition of a GIS which will meet their needs.

**Figure 1 - Four Phase GIS Implementation Process**


Phase I
System Specification
Step 1: User Requirements Definition
Step 2: Alternatives Analysis and Recommendation
Step 3: Develop Implementation Plan
Phase II
System Design
Step 1: Design System and Operating Procedures
Step 2: Produce Detailed Database Design
Step 3: Conduct Pilot Project (optional)
Phase III
System Acquisition
Step 1: Request and Evaluate System Proposals
Step 2: Benchmark Test Candidates
Step 3: Select Vendor(s) and Acquire System
Phase IV
System Commissioning
Step 1: Develop Organization, Assign and Train Staff
Step 2: Prepare Site Facilities/Install System
Step 3: Conduct Data Conversion
Step 4: Develop Applications
Step 5: Develop Operating Procedures
Step 6: Conduct Pilot Project (optional)
Step 7: Transition to Operational Mode

**1.2 Yukon department of renewable resources** Yukon DRR is charged with the responsibility of ensuring that " ... the environment and renewable resources of the Yukon are managed and used on a sustainable basis." The Department's objectives must balance a diverse range of requirements including:

- conservation of renewable resources and the protection of genetic diversity;
- support of development of the renewable resource economy in a sustainable manner;
- integration with other economic sectors to provide optimum benefit to all Yukoners;
- ensuring that environmental management reflects the rights and responsibilities of the Yukon's aboriginal peoples;
- ensuring the involvement of all Yukoners in the development and review of programmes, policies, legislation and regulations.

A more complete discussion of these issues can be found in "Yukon Conservation Strategy: Our Common Future" (Government of Yukon, 1989).

DRR has five Branches and ten Sections situated at three locations in Whitehorse, and Field Services staff throughout the Yukon. Approximately thirty of the Department's staff are potential users of the RLIS.

**1.3 Project team** The project team comprised three personnel: a Project Manager/GIS Consultant, a Systems Analyst, and an Applications Specialist. The Project Manager was responsible for: client liaison, running of user workshops, GIS issues, and content of the requirements analysis. The Systems Analyst was responsible for: selection of system diagramming methods, development of documentation standards, use of the CASE tool, and the development of the project deliverables. The Applications Specialist was responsible for: liaising with DRR personnel, gathering requirements information, checking diagrams for correctness, and the content of the requirements analysis.

The Project Manager and Systems Analyst were provided by Crippen Consultants. The Applications Specialist was a DRR staff member who had been closely involved in the Phase I planning process. In addition, an advisor from the Systems & Computing Services Department of the Yukon Government was also involved to ensure that the process followed system development guidelines.

## System diagramming methods

**2.1 The systems development lifecycle** Many public and private sector organizations use a systems development lifecycle approach to deliver applications programs to users (at the simplest level, one user application equals one problem solved). The lifecycle approach divides a program's existence into a number of stages starting with initiation, through to implementation, to maintenance, and finally to replacement.

The lifecycle approach provides managers and users with a common model that facilitates the planning and management of the development process. Although there are many examples of this methodology, most include the same essential components (refer again to Figure 1 for an example of this type of process).

**2.2 System diagramming Methodologies** Throughout the development process, system designers use various methodologies to assist in the analysis, design, programming, and testing of software applications. In the second phase of the system development process shown in Figure 1, the GIS implementation specialist is faced with the challenge of documenting and analyzing user requirements to produce a detailed system design. This need is met using system diagramming methodologies.

System diagrams represent user applications as graphical models. They are particularly helpful during detailed requirements analysis when users are trying to communicate abstract ideas to the system designer. There are many different methods that can be considered by the GIS analyst. Table 1 provides three examples of diagramming methods and their symbols.

**2.3 Data flow diagrams and data dictionaries** For this stage in the DRR project, Crippen decided to use Data Flow Diagrams (DFD's) and a Data Dictionary. Each is defined below.

*Data Flow Diagram* A DFD allows users to identify the system, its boundaries, and its interaction with its environment without requiring the identification of which functions are computerized and which are not. The DFD methodology used by Crippen has four symbols depicting: external entities, processes, data stores, and data flows. Table 2 presents the symbols, their meaning, and examples of their use. Usually, one DFD is drawn for each user application.

*Data Dictionary* A Data Dictionary can be defined as an organized reference to the data content of an organization's programs, systems, databases, files or manual records (Brathwaite, 1990). For natural resource mapping, a Data Dictionary is usually a typed glossary of map and attribute features. Rather than using a simple structure like the alphabetic order of an English language dictionary, a more versatile dictionary might also incorporate a hierarchical or relational structure to represent these features and the relationships that connect them. For example, a data dictionary of maps and their map features might have an alphabetized list of map names at its highest level, followed by ordered lists of features contained within each map at the next level. The dictionary might also track features which are common across a set of maps. Given the large number of applications and maps used by DRR, the need for a Data Dictionary to manage the names in the DFD's was a clear requirement (see Meador (1988) and Brathwaite (1990) for more detailed discussions of Data Dictionaries).

## Table 1 - Examples of System Diagramming Methodologies

| **Type of Symbol** | **Diagramming Method** | | |
|---|---|---|---|
| | Bubble Data Flow Diagram | Document Flowchart | Structure Chart |
| Operations | ○<br>transform | ○ ▭<br>operation inspection | ◇ decision ▭ process<br>▱ input/output |
| Connectors | →<br>✳ ⊕<br>AND OR | →<br>○ ⬠<br>on-page off-page | →<br>○ on-page ⬠ off-page<br>⟲ loop ◇ decision |
| Input/Output | shown in connector arrows | shown in terms of media | ●→ control data<br>○→ data to be transformed |
| Store | —<br>or<br>═ | ▭ ▽<br>document store | none |
| Source or Sink | ▭ | none | ( environment |
| System Boundary | – – – | none | none |

Adapted from E.N. Baylin (1988)

## Table 2 - Data Flow Diagram Symbols

| SYMBOL | DEFINITION | EXAMPLES |
|---|---|---|
| ▭ | External Entity: The external entity represents any outside object that influences the system. | Public, Indian Bands, Industry, Public Interest Groups, Government Agencies, Cabinet. |
| ▢ | Process: The process is used for any task, procedure, or activity that transforms data (whether a manual or automated process). | Receive Data, Interpret Airphotos, Verify Data, Update Map, Produce Sketch Map, Overlay Maps. |
| ⊏ | Data Store: The data store represents a place where data are stored or where data originate. | Planning Data, NTS Base Map, Protected Area Inventory, Land Use Permit Maps, Climate Data, Written Recommendation. |
| → | Data Flow: The data flow indicates the direction and name of the data that are being transported from one process, data store or external entity to another. | New Woodcutting Permit, Roads and Trails Location, Identified Conflicts, Natural Feature Location. |

## Analysis of user requirements using CASE tools

**3.1 Why use a CASE tool?** In parallel with the selection of the system diagramming methodology, Crippen also had to decide what kind of productivity tools would be used to develop the DFD's and the Data Dictionary. One possible alternative was to use a computer drafting tool such as MacDraw or PageMaker to draw the DFD's, and a spreadsheet or PC database package for the lists of names of maps, map features, external organizations, data flows, etc.

The key problem with this approach is that the diagrams and the dictionary are not connected. After an analyst has completed drafting a DFD, the names of map features, external entities and data flows associated with that diagram still have to be entered into the spreadsheet or database file. Additional problems with this approach become apparent once the users decide to add or change flows, names of external entities, names of maps, map features, etc. Clearly, there is a need for a tool which integrates the graphical power of a DFD with the usefulness of a structured data dictionary. CASE tools meet this need.

**3.2 Definition of CASE tools** "CASE [tools are] software that develops software" (Wylie, 1989). They are used to increase productivity in all aspects of the system lifecycle. In the front-end requirements analysis stage, the tools provide an interactive drafting environment which allows an analyst to draw the diagrams while the data dictionary is being maintained in the background. Most tools offer a number of diagramming methods including: DFD's, entity-relationship data models, structure charts, logical data models, etc. CASE tools are available for a variety of platforms and operating systems. Software prices range from $8,000 to $30,000 per analysis workstation for a usable set of modules.

There are a number of benefits that an organization can realize by making use of CASE tools including: risk management, cost reduction, increased flexibility, management of complexity, and improved responsiveness to users. There are also a number of drawbacks to CASE tools including: a need for specialist skills, the technology can alienate certain users, and additional project cost for the CASE license. The decision as to whether or not to use a CASE tool must balance hard costs such as software licenses with operational costs such as the time required for an analyst to update DFD's and the Data Dictionary by hand.

## Project description

**4.1 Overview** The DRR project was divided into five tasks:

1. Define methodology, develop instructions;
2. Run user workshop and gather data;
3. Gather detailed requirements data;
4. Analyze data using CASE tool;
5. Produce System Design document.

The elapsed time for completion of these tasks was four months. The main reason for the long time period was the fact that the participants had their regular daily responsibilities to attend to before they could make time to document their applications.

Each task is described below.

**4.1 Define methodology, develop instructions** Using the existing User Requirements and Implementation Plan documents from the Phase I project, the Project Manager and System Analyst selected the appropriate system diagramming methodology. The project team then developed a handout for use at the user workshop.

The decision to use a CASE tool was also verified at this step. Given the number and nature of the applications and databases, as well as the complexity of the organization, the use of a CASE tool was seen a mandatory requirement for this second phase.

**4.2 User workshop and data gathering** The user workshop was run by the Project Manager with assistance from the Applications Analyst. The purpose of the workshop meetings was twofold: to review the system development process, and to introduce the system diagramming methodology. The review of the development process was required because it gave the users some perspective on why they were in the meeting room and why their input was critically important at this stage.

The meetings were organized into small groups of users with a common set of applications. After discussing the diagramming methodology, the group worked through one or two examples on a whiteboard to show how the process would work. Once this step was complete, the users were prepared to develop their own diagrams.

**4.3 Gathering detailed requirements data** The users were tasked with documenting their applications over a two week period. This allowed them sufficient time to draft and re-draft their diagrams before submitting them to the Application Specialist. In addition, it was important to allow the users to work in their offices where they would have direct access to the maps, forms, tables, etc. that they would be naming in their DFD's. During this period, the Applications Specialist was very busy reviewing diagrams and answering questions.

Because an initial Data Dictionary was developed in Phase I of this project, it was possible to give the users a skeleton for adding names of map- and attribute-based features to the dictionary. To ensure that the diagrams were complete and followed standards, the system analyst travelled to Whitehorse to interview key users in person over a three-day period.

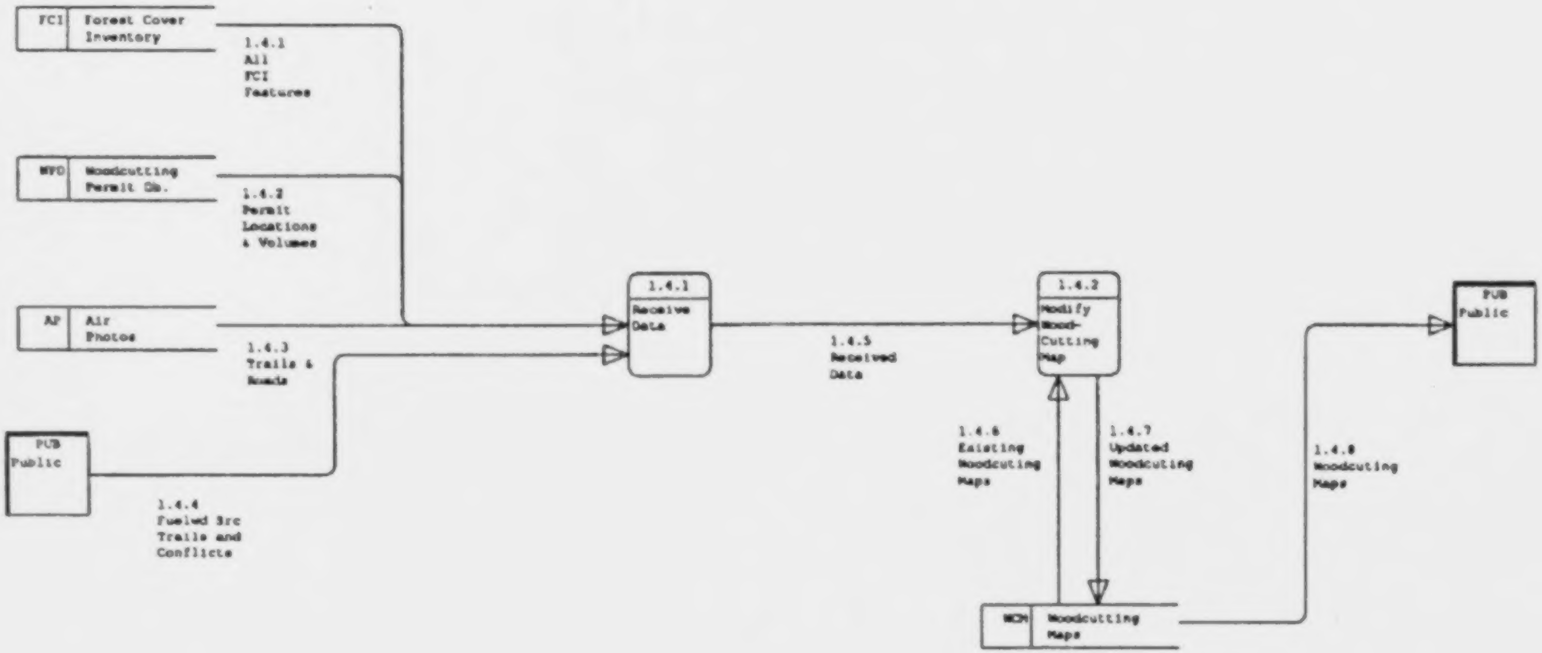

FCI
Forest Cover Inventory
1.4.1 All FCI Features
WPD
Woodcutting Permit Db.
1.4.2 Permit Locations & Volumes
AP
Air Photos
1.4.3 Trails & Roads
PUB
Public
1.4.4 Fueled Src Trails and Conflicts
1.4.1
Receive Data
1.4.5 Received Data
1.4.2
Modify Wood-Cutting Map
1.4.6 Existing Woodcuting Maps
1.4.7 Updated Woodcuting Maps
WCM
Woodcutting Maps
1.4.8 Woodcuting Maps
PUB
Public

Figure 2 - Example of a Data Flow Diagram

**4.4 Analysis using CASE tool** Once the draft DFD's were verified with the users, the System Analyst automated the diagrams using the CASE tool. Because the system analyst was closely involved with the selection of the diagramming methodology, the process of using the CASE tool was straightforward. After the diagrams were complete, they were returned to the users for verification and final changes. In some cases, the second and third version of a diagram did not at all resemble the first version.

**4.5 Produce system design document** Once the DFD's and Data Dictionary were signed off by DRR, the System Design document could be produced. This document is typically divided into five main sections:

1. Terms of reference/explanation of diagramming methods;
2. Business operations and computer systems analysis;
3. Global data architecture;
4. Detailed functional requirements;
5. Other criteria for vendor selection.

This process produced a complete version of Sections 1 and 2. The remaining sections are to be completed in a subsequent project.

Because the CASE tool produces automated summary listings and reports, the production of the document was a straightforward task of adding a table of contents and writing additional explanatory and narrative text.

**Table 3 - Sample List of Data Stores**

| Map/Database Type | | Map/Database Name | CASE Name |
|---|---|---|---|
| (Group Name) | (Series Name) | | |
| **Internal Maps/ Databases (IMD)** | Base Information (BI) | Outfitter Areas | OA |
| | | Game Management Areas | - |
| | | Conservation Officer Districts | - |
| | | Ecoregions/Natural Regions | ENR |
| | | Roads, Trails and Special Areas | RT |
| | Parks & Outdoor Recreation (P&OR) | Natural Features Inventory | NFI |
| | | Recreation Features Inventory | RFI |
| | | Canadian Heritage Rivers | - |
| | | Potential Recreation Site Inventory | PRSI |
| | | Campground/Recreation Site Map | CRSM |
| | | Candidate Parks Data | CPD |
| | Other Resources (OR) | Protected Areas Inventory | PAI |
| | | Woodcutting Maps | WCM |
| | | Existing Land Use Inventories | ELUI |
| | | Shoreline Inventories (new) | - |

**4.6 Examples of results** Twenty-six user applications were documented using this method. Figure 2 and Tables 3 through 7 provide examples of the types of summaries and reports produced using the CASE tool. The text included in the summary descriptions are drawn directly from the System Design Document. The model currently contains 100 data stores, 32 external entities, 111 process descriptions, and over 600 data flows.

The power of this documentation method becomes immediately apparent when one attempts to "read" a DFD: the Data Dictionary provides a complete, structured description of each symbol in the diagram, thus allowing a GIS Analyst to understand how the spatial data is being manipulated. Once the documentation process is complete, the project team will use this model to begin the process of developing a Global Data Architecture and the High- and Low-Level Functional designs.

**Table 4 - Sample List of External Entities**

| Organization Type | Organization Name | Branch or Specific Name | CASE Name |
|---|---|---|---|
| Public | Public | (any or all of below) | PUB |
| Industry | Industry | (any or all of below) | IND |
|  | Mining Company (ies) | - | - |
|  | Forest Company (ies) | - | - |
|  | Trapper (s) | - | - |
|  | Outfitter (s) | - | - |
| Government Agency | Government Agency | (any or all of below) | GA |
|  | Ag Canada | - | AC |
|  | DFO | - | DFO |
|  | DOE | AES | DOE-AES |
|  | DOE | CPS (formerly Parks Canada) | - |
|  | DOE | CWS | - |
|  | DIAND | Forestry | D-F |
|  | DIAND | Geology | D-G |
|  | DIAND | Lands | D-L |
|  | DIAND | Land Use | D-LU |
|  | DIAND | Land Claims | - |
|  | DIAND | Mining Recorder | D-MR |
|  | DIAND | Water Resources | D-W |

**Table 5 - Example of a Data Store Summary**

| CASE Name | Diagram Label | Data Store Description |
|---|---|---|
| AAD | Agriculture Applic. Dbase | Agriculture Application Database - internal database used to store data about agriculture and grazing land applications reviewed by the Department. Includes SAS attribute database and 1:30,000 mylar maps. |
| AP | Air Photos | Air Photos - photos generally purchased from the National Air Photo Library for specific projects and maintained by the departmental air photo library. Because of project specific nature, includes a wide variety of scales, types and quality. Some applications use photos borrowed from DIAND Forestry or Energy, Mines and Resources Canada. |
| AR | Agriculture Recommendation | Agriculture Recommendation - Branch recommendations on a specific LARC or agriculture/grazing application used to develop departmental recommentation. Written, with supporting map if applicable. |

**Table 6 - Example of an External Entity Summary**

| CASE Name | Diagram Label | External Entity Description |
|---|---|---|
| AC | Ag. Canada | Agriculture Canada - Federal Department responsible for acquisition and maintenance of soil databases in the Yukon. |
| COM | Committees | Committees - an entity grouping various committees whose communication with DRR involves spatial data. It includes the land application review committees, land use planning commission, and park management planning committee. |
| D-F | DIAND Forestry | Department of Indian Affairs and Northern Development: Forest Resources - federal agency responsible for management of forest resources in the Yukon. |
| IND | Industry | Industry - an external entity grouping the various industrial companies whose communications with DRR involve spatial data. Includes mining and forestry companies, trappers and outfitters. |

**Table 7 - Example of a Process Summary**

| CASE Name | Diagram Label | Process Description |
|---|---|---|
| 2 | Land Applic Review | Land Application Review / DAP |
| 2.1 | Parks Rec. to Land Apl Review | Parks Recommendation to Land Application Review Process |
| 2.1.1 | Receive Application | This process receives the Land Application from the external entity (usually Regional Planning or Fish and Wildlife). |
| 2.1.2 | Receive Data | This process gathers the necessary maps and data which will be overlayed. |
| 2.1.3 | Overlay Maps | At present, this process simply overlays the maps visually and identifies conflicts. In the future, it should also include such functions as: quantity (extent) of overlays, view of data all at once, visual impact analysis (i.e. from elevation model). |
| 2.1.4 | Develop Recommendations | Develop Recommendations for submission to committee. |
| 2.1.5 | File Application and Recommend | File Application and Recommendations. |

## Discussion

### Methodology issues

- The use of a diagramming methodology was attempted in Phase I of the development process without much success (Johnstone, 1991). This was largely because the users were documenting their GIS needs for the first time through a lengthy questionnaire: asking them to document their needs using DFD's and a Data Dictionary at the same time was not reasonable. In this second phase project, the use of DFD's was much more successful.

- The project team had to spend a lot of time with the users to answer questions and to ensure that they were properly applying the methodology. While this took more time than anticipated, it was worth the investment.

- A major benefit to this project was the fact that an initial user requirements analysis was performed as part of the implementation planning phase. The Phase I User Requirements document provided a base for the detailed work and allowed the project team to develop a solid initial data model before going to the users for additional detail. This was also very helpful in managing the scope of the work.

- In some cases, the applications identified by the users were theoretical (i.e. an application they would like to address if a GIS was in place). The team chose to allow the users to document these applications since it would help to develop a vision of how the RLIS might be used in the future. However, most of the detailed analysis effort will be focused on priority applications.

- System diagramming methods are most effective for documenting applications which have a highly structured or repetitive aspect (e.g. driver's license renewal, paying a bill through an Instant Teller). Since the applications being addressed by Parks, Resources and Regional Planning are "fuzzy" in nature, it may not be possible to state definitively that there is one and only one way to solve a particular problem. Many different versions of a single application could be developed to address nuances in species, geographic region, etc. Rather than documenting each and every version, our solution was to attempt to develop a generic application which captured the essence of the user's analysis process.

**Project team**

- A multidisciplinary team approach is mandatory for this type of work. Given the variety of disciplines that were used on this project, it would not be reasonable to expect that any one person could provide all of the skill sets.

- In order for the user workshops to be successful, the workshop leader should be trained in facilitating methods (i.e. brainstorming, group decision making, decision analysis, multivoting). In some organizations, this process is called a JAD (Joint Application Development) session.

- It was particularly important to have a system analyst on the project team. This staff position provided important input on applicable diagramming methods, development and maintenance of standards, and design of solutions and workarounds during the analysis stage.

- There are arguments for and against having a system analyst interview the users directly. While most analysts can run a good interviewing session, they do not understand the applications and can have difficulty in developing a rapport with the users. We believe that it is preferable to use an internal applications person, who combines an understanding of the user applications and the development process (assuming that type of person is available).

**CASE tools**

- Because the users were defining their requirements in detail for the first time, there was a need to iterate through some of the diagrams. While it is desirable to continue to improve a diagram, there is a danger of iterating too many times. The CASE tool makes it easier to do this because the cost of iterating is less than with a manual system.

- Because the CASE tool offers the capability of capturing large amounts of information, there was a risk of adding too much detail too soon. This was managed by focusing on a singe DFD per application. After completion of the current work, DRR will return to the diagrams and add further detail about the map features, the actual steps taken within each process, etc.

- The DRR project clearly demonstrated the benefits of using CASE as enabling tool for applying the system diagramming methodology. The benefits were most clear once the users began to iterate on their initial diagrams by changing flows, names, etc. If it were not for the flexibility of being able to easily re-draw eight or nine complex DFD's after two or three names have been changed, it would have taken much longer to respond to the users.

- CASE tools will become less and less the exclusive domain of speciality firms. As the cost of licenses progressively drop along the demand curve, more and more organizations will be able to justify the use of this type of tool. For many organizations (e.g. forest companies, municipalities), the most viable option at present is to subcontract the use either to an outside consultant who currently makes use of the tool or to an in-house MIS department. In fact, many GIS users will probably discover that their organization's MIS department already uses CASE tools.

- Given the current cost for a CASE license, this tool will only find cost-effective applicability for situations where there are: multiple users, complex data structures, sharing of data across disciplines, and a large number of applications. For less involved implementations, a cost-effective workaround might use a drafting package and a spreadsheet or PC database tool.

## References

Baylin, E.N. (1988) "System Diagramming Methodology" in *Handbook of MIS Management, 2nd Edition*, ed. R.E. Umbaugh, Auerbach Publishers, Boston, pp. 257-271.

Brathwaite, K.S. (1990) *Applications Development Using CASE Tools*, Academic Press Inc., San Diego, 263 pp.

Johnstone, W.M., B. Hawkings, L. Crooks (1991) *Development of a GIS Implementation Plan for the Yukon Department of Renewable Resources* GIS '91 Symposium, Vancouver, B.C., Feb. 12-15, pp. 253-263.

Meador, J.G. (1988) "Understanding Data Dictionary Technology" in *Handbook of MIS Management, 2nd Edition*, ed. R.E. Umbaugh, Auerbach Publishers, Boston, pp. 415-425.

Wylie, C. (1989) "CASE Tools for Systems Analysis, Design, Programming and Software Maintenance" in *Handbook of MIS Management, 1989-90 Yearbook*, ed. R.E. Umbaugh, Auerbach Publishers, Boston, pp. 107-118.

# D5 Data issues

**SAIF - conquering space and time**

H.A. Kucera et al., Ministry of Environment, Lands and Parks, Victoria, BC, Canada

**NAD27 to NAD83 conversion in British Columbia: National Transformation, Version 1.1**

Amin Kassam, Ministry of Environment, Lands and Parks, Victoria, BC, Canada

**Putting the records in order: GIS database construction**

C. Warren Ferguson, Cantex Conversion Services, Kamloops, BC, Canada

**Technology trends for GIS applications**

Michael J. Fisher, Data General International, Westboro, MA, USA

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# SAIF — Conquering space and time

**H.A. Kucera**
**Mark Sondheim**

Geomatics Unit
Surveys and Resource Mapping Branch
Ministry of Parks and Lands
4th Floor, 1802 Douglas Street
Victoria, B.C. V8V 1X4

## Abstract

*The principle drawback in the use of existing Geographic Information systems for resource management has been the inability to represent multidimensional or multitemporal data. To conduct spatial analysis in higher dimensions three major issues must be resolved.*

*First, we must distinguish between multiple occurrences of similar objects or the repeated occurrence of a single object along a sampling path i.e. the multiple intersections of a borehole with folded strata. Secondly, since many phenomena are not visible to the researcher we have imperfect knowledge of the objects under investigation. The orientation of the objects cannot be assumed and the continuity and orientation of the boundaries of objects cannot be discretely quantified. In many cases the objects are modelled quantities, therefore statistical error must be identified and associated with each object. Lastly we must recognize the dynamic character of natural phenomena. The inherent variation through time requires repeated sampling of objects or phenomena which generates multiple data sets that are uniquely identified only by the time of sampling.*

*The data model underlying the Spatial Archive and Interchange Format (SAIF) provides the architecture required to comprehensively represent natural phenomena. A discussion regarding how the above issues are addressed using the SAIF architecture will illustrate the significant advantages of adopting the format.*

## Introduction

One of the major drawbacks of existing Geographic Information systems for use in natural resource management has been its inability to represent or analyze multi-dimensional data. In traditional GIS applications using conventional database structures it is difficult to distinguish between multiple occurrences of similar objects or the repeated occurrence of a single object along a sampling path i.e. folded or faulted strata.

Success in conducting spatial analysis in higher dimensions requires an underlying data model than can provide the following:

- Support for 3 independent coordinate axes (X,Y,Z) plus providing attribution at each independent location in space.
- True 3-D functionality (the value of Z is not a function of X and Y) such that each X,Y pair can have multiple Z components.
- Support for geostatistical operations (accurate error capture and representation).
- Support queries to a Data Base Management System (interactive query and update, real-time analysis).
- An expression of the time component that is flexible enough to allow multitemporal analysis, monitoring, navigation and forecasting.

**Background**

As of June 1991, Version 1 of the British Columbia Spatial Archive and Interchange Format (SAIF) developed by the B.C. Ministry of Lands and Parks was accepted by the Canadian General Standards Board (CGSB) Committee on Geomatics as the draft Canadian Standard for the exchange of Geomatics Information. Version 2 of SAIF, completed in December 1991, is currently being used by MacDonald Dettwiler and Associates in the development of B.C's Land Information Infrastructure (LII).

SAIF, a powerful and expressive object oriented data model for Geomatics Information Exchange, has been undergoing development since 1989. This data model, together with its associated schema definition, provides the structures for the description of geometry, relationships and attributes required for the complete and unambiguous transferal of geographic information from one platform to the next. SAIF is also extremely flexible, allowing geometry to be expressed in vector and raster structures in up to four dimensions, with or without associated uncertainty. Temporal issues are addressed by making time available as a coordinate, a time stamp or a duration. SAIF supports a number of different relationships including spatial and temporal topology. This flexibility supplies the capability to address true three dimensional and multi-temporal data analysis.

## The formal SAIF specification: the data model and schema

The creation of the SAIF Data Model and Schema required a language and methodology. To construct the model, a technique was selected that allowed separation of specification from the implementation of the data model and that supplied primitive concepts and elemental components for modelling.

The SAIF specification is described using a language capable of expressing object-oriented concepts. The specification of SAIF has two parts: SPECTalk and SAIFTalk. The first part is a formal description language used to describe the data model. The second part is the extension of the formal language used to define specific geomatics concepts. In general terms SPECTalk is the definition of the Data Model and SAIFTalk provides the Schema Definition. The overall structure of SAIF is described in terms of ***Abstract Objects, Enumerations*** and ***Primitives*** as shown in Figure 1.

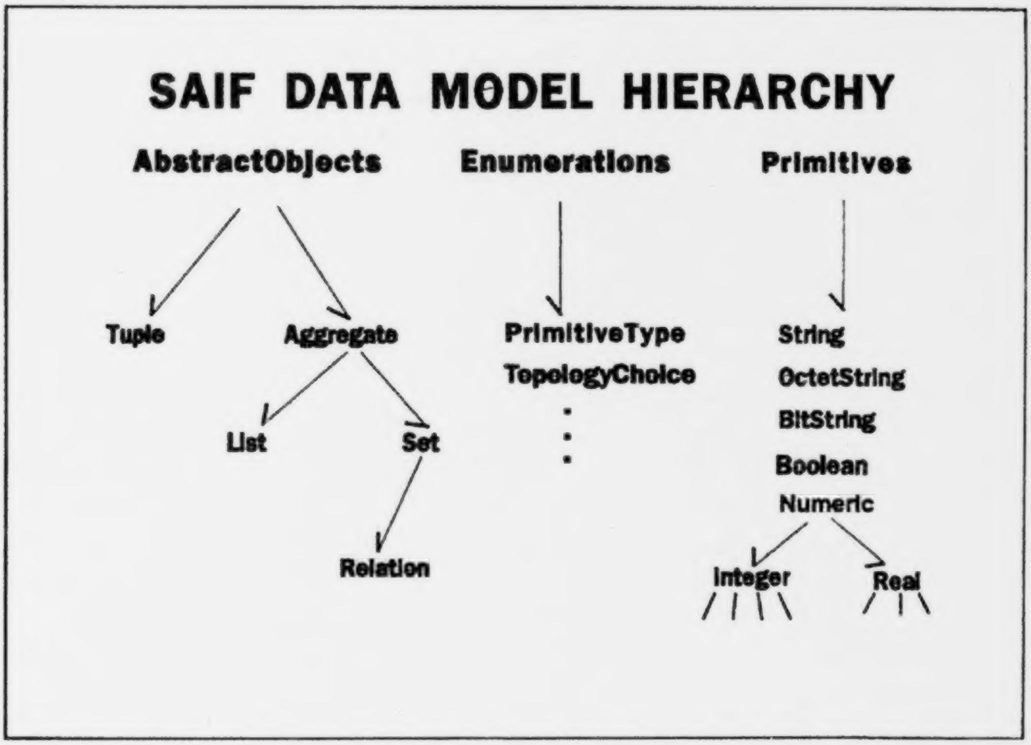


**Figure 1**: *The Data Model hierarchy illustrating the basic building blocks of SAIF.*

All other classes which represent the field of geomatics are constructed from these three primitive classes. The basic building block in SAIF is the Abstract Object which is the abstract super class to which all other objects belong (excluding primitives and enumerations). As subclasses of Abstract Object, each object in the hierarchy inherits a unique internal identifer.

***Abstract Object*** is at the highest level in the Data Model Hierarchy. *Tuple, Aggregate, Set, List and Relation* are subclasses of Abstract Object The most frequently used subclass is ***Tuple,*** which is an ordered set of unique attributes that can serve to define the majority of SAIF structures.

***Enumerations*** are a discrete set of categorical choices which can be used to provide values for a particular data type. For example, the data type Eye Colour may have specific valid values, such as blue, green, brown, etc. ***Primitives*** are the atomic objects from which all other objects are constructed.

### Object orientation

The SAIF specification relies heavily on concepts from the Object Oriented Data Modelling paradigm (Coad and Yourdon, 1990). The foundations of SAIF are the basic concepts *class, object and inheritance.* A *class* is a named abstraction for a particular type of entity or thing. *Objects* are specific instances of classes. Objects in a given class are recognized as instances of a particular type of entity. The "character" of the object is given by the attributes associated with the class.

Classes are related to one another through the structure of *inheritance* hierarchies. One class may be a generalization of another class or it may be a subclass of a more general class. A subclass inherits attributes and properties from the parent class and may add additional attributes which serve to distinguish the subclass from its ancestors. The concept of inheritance is discussed in detail in the section on the Coordinate superclass and its subclasses.

The depth of description, flexibility and extensibility of SAIF is possible through the use of hierarchy, inheritance and encapsulation. Encapsulation is the term used to describe the concept of "Information Hiding". By applying encapsulation to the overall SAIF system, each component of the system is encapsulated or hidden from other components. The encapsulation can include the behavior of objects and methods associated with objects. This allows components of the system to continue to evolve without resulting in changes to other aspects of the system. The isolation of potentially volatile components of the system is seen as essential to the long term success of the standardization effort (Coad and Yourdon, 1990).

## SAIF's contribution to three dimensional GIS

The difficulties encountered by conventional GIS implementations dealing with three dimensional data and multi-temporal analysis are due in part to the limitations of the Relational Data Model (Ris, 1991). The Object Oriented Data Model underlying the SAIF architecture provides the structures required to represent and analyze natural phenomena. A discussion regarding how the above issues are addressed using the SAIF architecture will illustrate the significant advantages of adopting the proposed format.

### Geographic objects

One approach to the problem of modelling in three dimensions is to treat each phenomena as a single geographic entity or "Object". Applying this approach allows us to locate any feature that has spatial limits or boundaries that can be measured or estimated. This concept is the basis of SAIF's approach to modelling. Each ***geographic Object*** in the SAIF Data Model has five components as shown in Figure 2.

The representation of geographic data in 3-D space is addressed in both raster and vector forms. SAIF provides for 3-D georeferencing of a geographic Object in both forms in three different ways; 1) using direct reference (coordinates), 2) indirect reference (indexing and gridding), 3) metric and nonmetric spatial relationships (such as bearing and distance or topology). The general structure of a geographic Object is shown in SAIF notation below. The notation includes not only a list of the valid attributes or components that comprise the Geographic Object but also the "domains" or classes from which these attributes come.

**Tuple**

| | |
|---|---|
| **subclass:** | **GeographicObject** |
| attributes: | [geometry] [relationships]<br>[metadata] |
| domains: | GeometricObject<br>Set (Relationship) Metadata |
| class attributes: | [feature type] |
| domains: | String |

This subclass of tuple, contains geometry, relationships, metadata and feature type The class may be used to carry data directly. More typically, subclasses of geographic object are defined by the user. subclassing is required in order to specify features, additional attributes, defaults, and/or additional restrictions and constraints. The definitions for the subclasses are provided in the metadata. *Note - that the square brackets [] imply that the attribute is optional.*

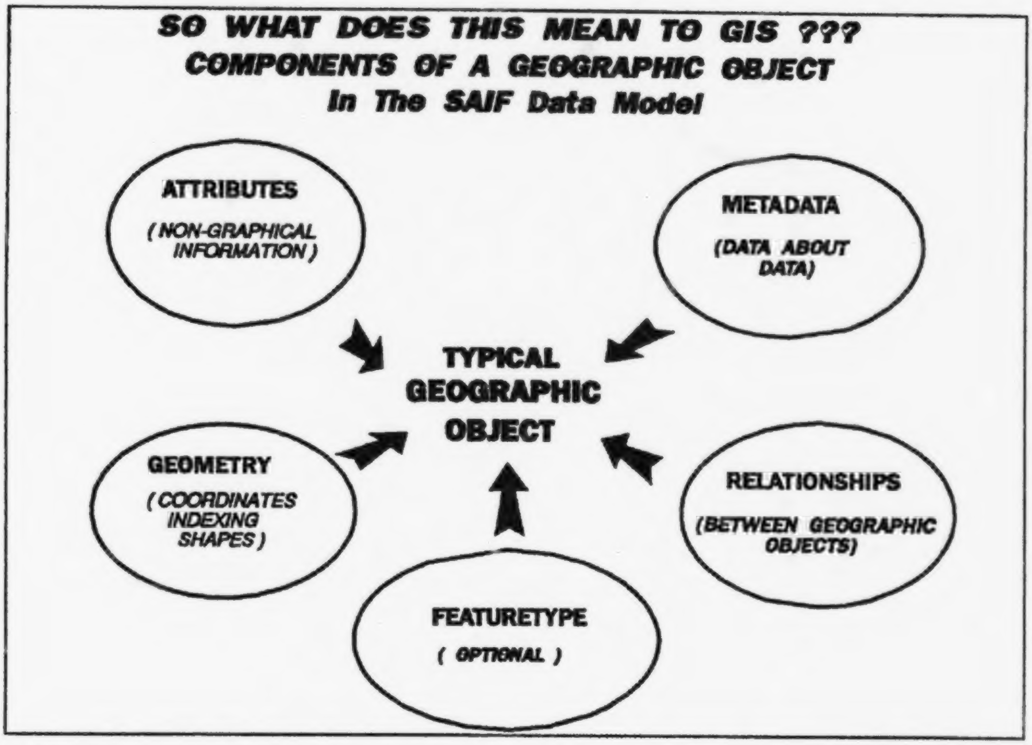

SO WHAT DOES THIS MEAN TO GIS ???
COMPONENTS OF A GEOGRAPHIC OBJECT
In The SAIF Data Model
ATTRIBUTES
( NON-GRAPHICAL INFORMATION )
METADATA
(DATA ABOUT DATA)
TYPICAL GEOGRAPHIC OBJECT
GEOMETRY
( COORDINATES INDEXING SHAPES )
RELATIONSHIPS
(BETWEEN GEOGRAPHIC OBJECTS)
FEATURETYPE
( OPTIONAL )

**Figure 2:** *The components of a GeographicObject in SAIF illustrating the richness of description available for any given entity.*

Geometric objects only become geographic objects when combined with metadata, relationships, attributes or feature types, which provide "meaning" to the geometry.

The classes Coordinate, index scheme and Relationship are subclasses of tuples that provide exact or relative spatial information associated with a geographic object. coordinate and indexing are imbedded in geographic objects as components of the geometric object, whereas relationships between geographic entities are specified at the higher geographic object. level (Figure 2). By limiting the georeferencing to these three data types, changes in resolution, units, and projection can be isolated from the definition of the geographic object. These three classes, together with their associated subclasses include *all* of the valid options of referencing spatial information in SAIF, however one or all of these objects may be omitted from a Geographic object definition. This implies for example that a) topological relationships are *not required*, and b) coordinate information is not required if, at some level, other relationships are used to identify the spatial reference for the object.

For a user or procurer of information for a specific system the method of georeferencing must be given meaning. The information resident in a database or being exchanged through a transfer must be rigorously defined by the metadata. The metadata includes the information required to produce physical representations for all variables used in the description of SAIF. Each entity described in SAIF is given meaning through the inclusion of Metadata in the geographic object.

## Coordinates

Part of the hierarchy for Coordinate and its associated subclasses is given in Figure 3. This diagram illustrates the basic concepts of inheritance. Note that SuperClass, Class and SubClass are relative terms and each simply refers to a given level in the hierarchy with respect to another class. In addition, the term "*abstract Superclass*" refers to a high level class within which attributes cannot be instantiated. The hierarchy and inheritance of attributes is indicated by the arrows and the indentation.

Further subclasses of Coordinate include CoordGeoid, CoordGeoidWithError and CoordPhotoStation. Note in Figure 3 that the Z and Time components are not a function of any other attribute. As any new elevation or time is specified, a new geographic object is created that is independent of any other object at the same X and Y coordinates. Note also that *fuzzy boundaries* can be defined if

## AN EXAMPLE OF THE SAIF CLASS HIERARCHY AND INHERITANCE OF ATTRIBUTES

**AbstractObject : (AbstractSuperClass)**

AbstractObject is an Abstract Superclass to which all other objects belong, excluding primitives and enumerations. Each Subclass of AbstractObject inherits a unique internal identifier.

- **Coordinate : (A Tuple)**

  | Attributes: | X | Y |
  |---|---|---|
  | Domains: | Numeric | Numeric |

  This class of objects is used for the allocation of position to objects in 2-dimensional space. It is a subclass of Tuple which is a subclass or "Type" of AbstractObject.

  - **CoordXYZ: (SubClass)**

    | Attributes: | Z |
    |---|---|
    | Domains: | Numeric |

    This subclass of Coordinate provides for the allocation of position to objects in 3 dimensional space. Note: Codinates in X and Y are inherited from the parent class of Coordinate.

    - **CoordXYZWithError: (SubClass)**

      | Attributes: | varX | varY | varZ | [covXY] | [covXZ] | [covYZ] |
      |---|---|---|---|---|---|---|
      | Domains: | Real | Real | Real | Real | Real | Real |

      This subclass allows for the specification of error related to a given position in 3 dimensional space. Note that square brackets indicate an optional attribute.

  - **CoordXYT: (SubClass)**

    | Attributes: | T |
    |---|---|
    | Domains: | Time |

    This subclass of Coordinate allows for the inclusion of time as part of a coordinate, rather than as an attribute. Time refers to the time of occurence or observation. Note: Codinates in X and Y are inherited from the parent class of Coordinate.

    - **CoordXYTWithError: (SubClass)**

      | Attributes: | varX | varY | [covXY] | [ varT] |
      |---|---|---|---|---|
      | Domains: | Real | Real | Real | Real |

      This subclass allows for the specification of error related to a given position and time of occurence. Note that square brackets indicate an optional attribute.

    - **CoordXYZT: (SubClass)**

      | Attributes: | Z |
      |---|---|
      | Domains: | Numeric |

      CoordXYZT incorporates time as a coordinate for multitemporal, 3-Dimensional spatial analysis.

      - **CoordXYZTWithError: (SubClass)**

        | Attributes: | varX | varY | varZ | [covXY] | [covXZ] | [covYZ] | [varT] |
        |---|---|---|---|---|---|---|---|
        | Domains: | Real | Real | Real | Real | Real | Real | Real |

        This subclass allows for the specification of associated error related to position in 3 dimensions and time.

**Figure 3:** *The structure of the Coordinate Superclass illustrating the inheritance of attributes.*

CoordXYwithError is used. Similarly, normal and fuzzy boundaries in three dimensional space can be specified using CoordXYZ or CoordXYZwithError, respectively. The same applies to CoordXYT and CoordXYTwithError, as well as CoordXYZT and CoordXYZTwithError.

The use of SAIF referencing methods will be illustrated by discussing subclasses of geometric object as shown in Figure 4. Examples regarding georeferencing of different types of Geometric objects will show the expressive power of the SAIF data model and schema.

## Time

In order to be a truly Geographical Data Model SAIF must be able to deal with the dynamic character of natural phenomena such as aquifers, ocean currents, migration paths, and growth patterns. The SAIF model provides the structures for storage and retrieval of the source data, as well as information derived from the source data. Additionally, as a basis for a temporal database, SAIF must provide the structures that will allow time related queries.

**subclass:** **TimeClass**

TimeClass is an abstract superclass used to define a variety of different concepts of time as subclasses. All uses of time in SAIF fall under TimeClass.

**TimeClass**

**subclass:** **General Time**

attributes: [second] [minute] [hour] [day] [month] [year] [day of week] [day of year]

domains: Integer Integer Integer Integer Interger Integer

General Time has no reference to a zero value. Normal concepts of date and time are handled by the subclass Time below.

**General Time**

**subclass:** **Time**

Time is referenced to the Gregorian calendar and Greenwich Mean Time. If instantiated at this level in of SAIF hierarchy, the first 3 attributes give time on a 24 hour clock, with second, minute, and hour. At lower levels, second is any non-negative integer. Other attributes are: day, month, year, day of week and day of year.

**Time**

**subclass:** **ANSIcTime**

defaults: minute:0 hour:0 month:0 year:70

This gives the current time, in seconds, elapsed since 00:00:00 GMT, January 1, 1970. It is equivalent to time as specified in ANSI C.

**ANSIcTime**

**subclass:** **Tiny Time**

attributes: nanosecond

domains: Integer

This notion of time is useful where extreme precision is required, (i.e. GPS or satellite imagery).

**TimeClass**

**subclass:** **Duration**

attributes: startTime endTime

domains: Time Class Time Class

constraints: startTime: Time or Fuzzy Time endTime: Time or FuzzyTime

The startTime and endTime of an object may be stated under geometryExtended.

**TimeClass**

**subclass:** **FuzzyTime**

attributes: time varT

domains: Time Real

If there is some uncertainty associated with a given time, this subclass may be used. varT represents the variance of time, with reference to the smallest time units used under *time*, unless indicated otherwise in the metadata.

The primary structure in which time is stored in the SAIF data model is Coordinate (Figure 3), however there are several other data structures which incorporate time. All valid uses of time are defined by the tuple, ***TimeClass*** or one of its **subclasses.**

SAIF allows values for time to be associated with an object in three different ways. Figure 3 illustrates that the geometry of an object may be defined by coordinate values using a subclass of Coordinate which includes *time* (such as CoordXYT). Alternatively, the attribute *timeStamp* may be used under GeometryExtended. Both of these methods allow the designation of a single value for time, indicating the time of occurrence or observation of the event. TimeStamp is the preferred approach when dealing with something such as an oil spill where the user is interested in a snapshot in time of a large complex dynamic object. The definition of the oil spill may in-volve numerous polygons and it would be inappropriate to attach time to each point or arc. The use of the *timeStamp* allows a time to be attached to the complete geographic object. This facilitates versioning and multi-temporal monitoring of phenomena.

There is a third option for association of duration to an object or event in SAIF. The event may be defined using startTime and endTime within the attribute *duration* under GeometryExtended. Duration is very

| | |
|---|---|
| **subclass:** | **GeometryExtended** |
| attributes: | object extension |
| domains: | GeometricObject Tuple |

This subclass allows a limited degree of attribution to be assigned directly to a geometric object. These attributes compose attributes that may be assigned to describe time:

| | | |
|---|---|---|
| timeStamp | from the domain: | Time or FuzzyTime |
| duration | " | Duration |

***timeStamp:*** In many situations it may be more convenient to add time to a geometric object in this fashion instead of using a coordinate with time. The value for timeStamp is the time of occurrence or observation.

***duration:*** The startTime and endTime of an object may be defined using duration.

## Geometry

### HIERARCHY OF GEOMETRIC OBJECTS

**GeometricObjects**

**PointClass**

**An Abstract superclass for defining the position of zero dimensional objects residing in higher dimensional space. The spatial dimension is indicated by the type of Coordinate.**

**LineClass**

**An Abstract superclass for representing 1-Dimensional GeometricObjects in higher dimensional space as defined through Coordinates.**

**AreaClass**

**An Abstract superclass for description of 2-D GeometricObjects in both raster and vector formats. The dimensionality of space in which the area exists is defined by the choice of coordinates. Note - It cannot be less than 2-D.**

**VolumeClass**

**An Abstract superclass for the definition of vector and cellular 3-D Geometric Objects. Cellular objects may include rectangles, spheres and ellipsoids.**

**RasterClass**

**An Abstract superclass for the definition of regular gridded structures (points, pixels, voxels) which have associated values. Position in space is defined by a combination of indexing and Gridding.**

**GraphClass**

**An Abstract superclass under which a variety of graph and network structures are defined.**

**TextClass**

**An Abstract superclass for the orientation and placement of independent textual information.**

**Figure 4:** *The subclasses of GeometricObject showing the usage of each subclass.*

useful when comparing two event sequences and can be used in conjunction with TemporalTopology and FuzzyTime. A SAIF definition of timeStamp and duration is given below:

### PointClass

A list of coordinates is the most common method of containing spatial information, particularly for vector-based geometric structures. As stated previously, the Coordinates are imbedded as attributes in the Geometric Object. There are several objects which are subclasses of the GeometricObject class (Figure 4), and each must have a defined position. The following example shows how position is defined for the PointClass within the data model structure.

**GeometricObject**

**Subclass:** **PointClass**

PointClass is an abstract superclass appropriate for the definition of the position of zero dimensional objects residing in the higher dimensional space.

**PointClass**

**subclass:** **Point**
attribute: position
domains: Coordinate

Point is a subclass of PointClass used for the position of single points, including nodes, origins of grids, placement of text, etc.

In the example, Point acquires a position from the domain Coordinate. Coordinate is simply a position with regard to a referencing system. Most subclasses of GeometricObject have their position and orientation in space defined through the use of the subclass PointClass, which carries position as attribute from the domain of Coordinate. Some of the possible subclasses of PointClass include: DEMPoint, CorrelationPoint, CorrelationPoint 3D, PointGrid, PointGrid3D and PointsInPointGrid3D. Through the use of these subclasses it is possible to define the position of two and three dimensional objects in multidimensional space using zero dimensional objects (Points). This class can be used for the description of discreet regular sampling locations, as would be taken in a soil survey or exploration drilling program. Some of the possible constructs of PointClass are shown in figure 5.

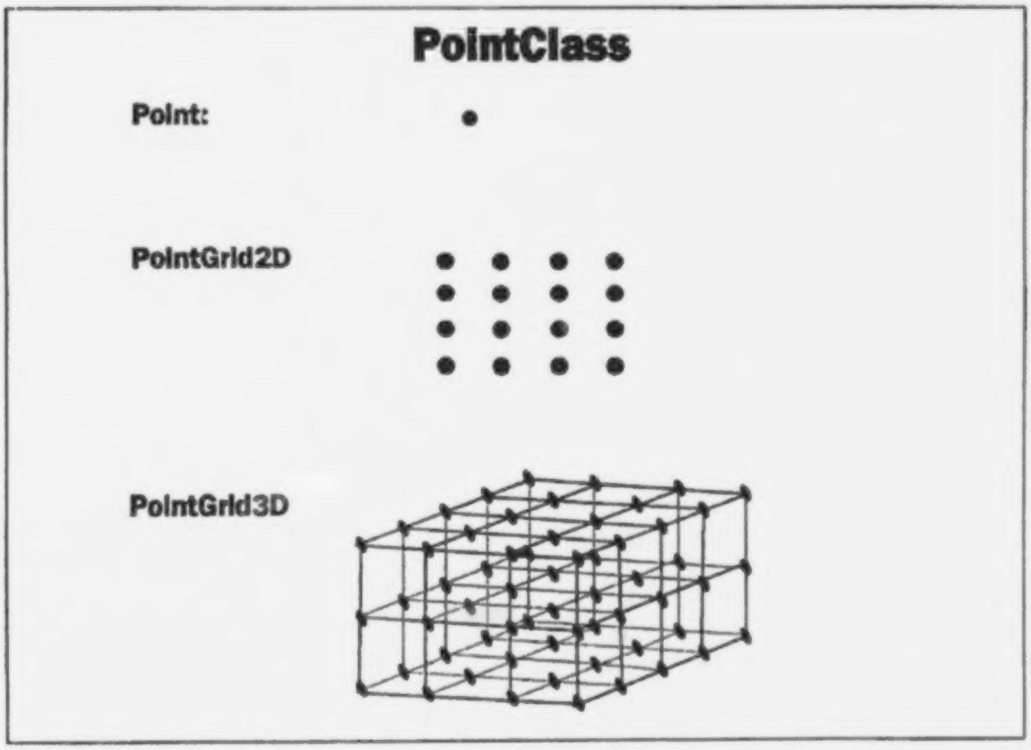


**Figure 5:** *Some of the constructs in the PointClass hierarchy.*

**LineClass**

LineClass builds on the structures provided by Coordinates and PointClass. The structure and type of line are further defined with a description and a structure that must conform to certain rules or ***constraints***. An example of the structure used to define an arc (note that the structure in this case is provided by a subclass called ArcDefinition):

**LineClass**

| | |
|---|---|
| **subclass:** | **Arc** |
| attributes: | position [description] |
| domains: | List(Point) ArcDefinition |
| defaults: | Description.type: line |

Arc is the locus of points connecting a list of points. Every Arc must include at least two points in the point list defining position, as well as a description which defines the mathematical method used for generating the locus. The square brackets enclosing description indicate it is an optional attribute. If the description is unspecified, the points on the arc are assumed to be connected by straight lines; that is, the type of arc (in ArcDefinition form the enumeration CurveChoice) is assumed to be *line*.

A variety of Lineclass subclasses can be defined that make use of multiple points, time, and direction. A partial list of these classes includes: ArcDirected, Path, Ring and Boundary.

**AreaClass**

AreaClass builds on structures provided by the SAIF hierarchy and components provided by objects with a lower dimensionality. Position is specified through the use of PointClass (which has a zero dimensionality); the x and y dimensions of the cells are provided by a subclass of the Numeric primitive (as is illustrated in the following SAIF formal definition of AreaClass) and may or may not match the grid spacing.

**AreaClass**

| | |
|---|---|
| **subclass:** | **Cell** |
| attributes: | [shape] positionInCell<br>[xDimension] [yDimension] |
| domains: | CellDefinition CellReference<br>Real Real |

Cell is an abstract superclass. With its subclasses, it defines different two dimensional cellular structures. The cellDefinition may be square, rectangular, circular, or elliptical. The positionInCell refers to where in the cell a point position refers; there are five choices, the centre plus the four corners of the cell. The xDimension and yDimension have units as specified in the metadata for the x and y coordinates, respectively.

**Cell**

| | |
|---|---|
| **subclass:** | **CellGrid** |
| attributes: | position |
| domains: | PointGrid |

This subclass of Cell provides for the specification of a grid of cells. Note that the starting location is defined by the choice for origin in PointGrid. All cells in the cellGrid are assumed to be of interest; the number of cells is the product of nRow multiplied by nColumn, (nRow and nColumn are attributes which come from PointGrid).

This example shows how Cellgrid makes use of the information previously defined in PointGrid which is a subclass of PointClass. There are a variety of other subclasses of AreaClass that are appropriate for the description of polygons, such as ComplexPolygons, Triangles, Triangular Irregular Networks (TINS), TINenvelope and mathematically defined surfaces. All of these subclasses are useful for defining commonly occuring entities in geographic space.

**VolumeClass**

Through the combination of object classes and the inheritance of attributes, it is possible to define enclosing surfaces as well as the volumes they enclose. An example of this would be 3-D Triangular Irregular Networks (Figure 5) and Complex Triangular Irregular Networks (TIN3D and TIN3DComplex). In this case, we use 2-D objects (surfaces) oriented in 3-D to define an enclosed 3-D volume. This structure is useful in the description of ore bodies or the description of drifts, shafts and stopes in an underground mine. The SAIF description for TIN3D structure is as follows:

**GeometricObject**

| | |
|---|---|
| **subclass:** | **VolumeClass** |

An abstract superclass for the description of 3-D Geometric objects. The dimensionality of the space in which they reside, 3 or higher, is defined by the choice of coordinates.

**VolumeClass**

| | |
|---|---|
| **subclass:** | **TIN3D** |
| attributes: | boundary |
| domains: | ***TINenvelope*** |

This subclass describes a volume bounded by a triangular irregular network. TINenvelope is a 2-D dimensional bounding surface used to define the limits of the volume. ***TINenvelope*** is a subclass of TIN which is a subclass of AreaClass. If a TIN3D has voids or holes in it, it may be modelled using the subclass TIN3DComplex. Each void is then defined by an inner boundary which is defined by an additional TINenvelope.The surface defined by the TINenvelope is multivalued (having more than a single z value for any x,y location) except for points on the periphery.

The GeometricObject defined above is an example of the descriptive capability of SAIF in a vector format. Several other 3-D structures such as Cell3D, Cell3DGrid, Box and BoxGrid can also be defined. Some Graphical examples of 3-D geometric objects are shown in Figure 6.

**RasterClass**

The gridded or cellular structures discussed above do not provide the means of carrying values directly. SAIF addresses the idea of values residing in these structures through the RasterClass hierarchy. The formal definition of the RasterClass follows.

**GeometricObject**

| | |
|---|---|
| **subclass:** | **RasterClass** |
| attributes: | values index |
| domains: | MeasuredValues IndexScheme |

This is an abstract superclass. With its subclasses, it defines regular, gridded structures consisting of points, pixels or voxels which have associated *values*.

The allocation of values to particular locations in a raster is controlled by:

i the particular subclass of IndexScheme used

ii the values of the components contained by *index*,

iii the particular subclass of MeasuredValues used

iv the values of the components contained by *values*.

*The* ***essential*** *difference between gridded and raster data is that Raster data is not only organized in a regular grid, but also incorporates the use of indexing to reference the data* ***values*** *directly to the structure of the GeometricObject.*

*Position (from grid) + Index + Values => Raster*

Four major subclasses of GeometricObject: PointClass, AreaClass, VolumeClass and RasterClass, can be used in conjunction with the spatial indexing schemes to provide unambiguous definitions of gridded values in a raster. There are also four different types of indexing methods that can be used to associate values with positions in the raster. The formal definition of the superclass, IndexScheme is as follows:

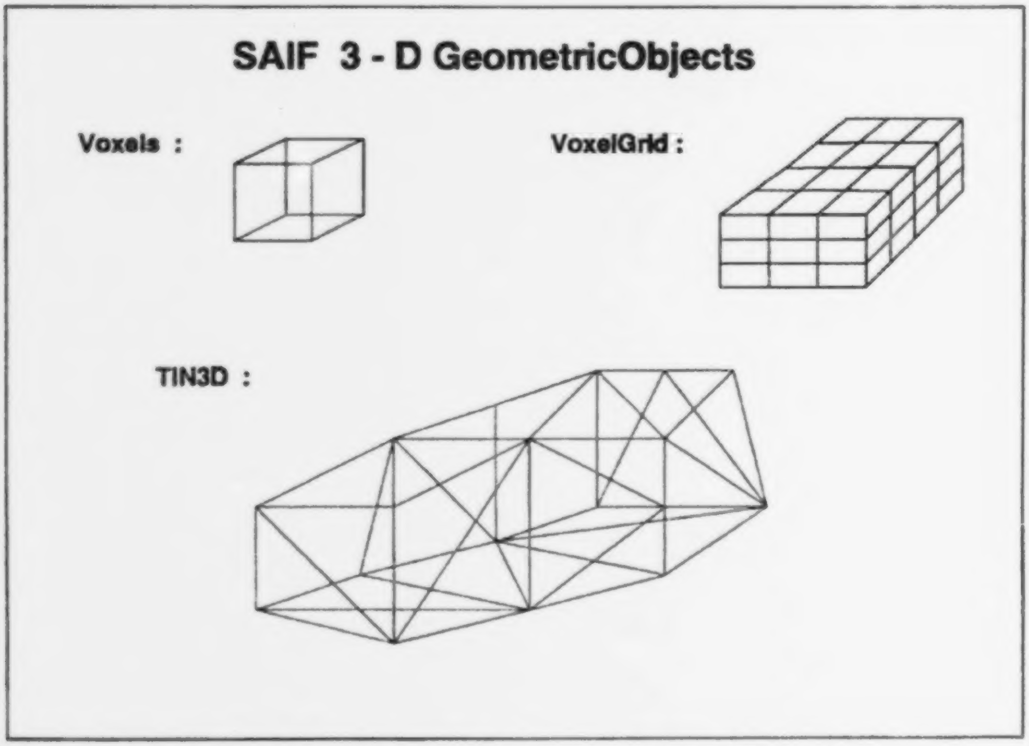


**Figure 6:** *Some of the 3-D GeometricObjects that can be described in the SAIF format.*

| **Positional Source** | **RasterClass Defined** | **Index Options** |
|---|---|---|
| PointGrid | PointRaster2D | Quadtree, RLE, OneToOne |
| PointGrid3D | PointRaster3D | Octree, OneToOne |
| CellGrid | PixelGrid | Quadtree, RLE, OneToOneCell |
| Grid3D | VoxelGrid | Octree, OneToOne |
| BoxGrid | BlockGrid | Quadtree, RLE, OneToOne |

By using the IndexClasses to provide the data storage format in conjunction with subclasses of VolumeClass, which provide position and orientation in space, we can tightly define any type of rasterized Geographic object.

**Tuple**

**subclass:** **IndexScheme**
attribute:
domains:

IndexScheme is an abstract superclass within which the possible schemes for referencing gridded data are defined. Four types of Indexschemes are possible OneToOne, RunLengthEncoded (RLE), Quadtree and Octree.

If index is of type Run Length Encoded (RLE), Quadtree or Octree, data compression is being applied and adjacent locations with identical *values* are grouped together. Digital imagery and typical GIS raster data can be housed in the structures under this subclass. By separating RasterClass from the frameworks provided by the purelygeometric classes *PointGrid, CellGrid, etc.*, sharing of these latter classes with different images occupying the same space is possible. The following table indicates how the components of SAIF can be combined to define a Raster.

### Further geometric objects

Other subclasses of GeometricObject exist for the description of graphs, networks, labels and text. These subclasses use the same general structures for referencing with regard to space and time that have been discussed. *How* these structures can be applied with regard to these classes is beyond the scope of this paper.

## Relationships

Relationship is defined by a Tuple which supplies the related geographic objects, the name of the relationship and any data required by the relationship. SAIF makes use of several relationships under the general framework given as follows.

**Tuple**

**subclass:** **Relationship**
attributes: RelatedObjects
domains: Set(Geographic object)

This is an abstract superclass under which the spatial and temporal relationships of different geographic objects may be specified.(Note - the primary geographic object to which the above objects are related is not included in the set.

The Set construction approach was chosen for SAIF because it allows for the designation of multiple related geographic objects having the same relationship to the primary object, without requiring multiple tuples. Some of the relationships defined in SAIF are: Neighbour, Compass, ConnectedTo, DistanceAndAngle, DerivedFrom, RelativeLinearDistance and UserDefined Relationship.

## Multi-dimensional topology

Topology is an example of the Relationship component of a Geographic object. The structures for describing topology can be adapted from the basic elements of describing any binary relationship. Examples of the relationships Topology, TopologyExtended, Temporal Topology and TemporalOffset are provided so that the reader can appreciate the structures for describing multi-dimensional relationships.

**Relationship**

**subclass:** **Topology**
attributes: [inclusion] [adjacency]
domains: InclusionStatus AdjacencyStatus

Topological comparisons require a clear understanding of the terms *boundary*, *interior adjacency* and *Inclusion.*

***boundary*** is defined as the union of the limiting case of all such neighbourhoods as their radii approach a length of zero. (modified from Egenhofer and Franzosa, 1991).

***interior*** is the set of points defining the object which are not on the object's boundary.

***Inclusion*** If the interiors of two objects share a subset of points in common, then their intersection may be said to be not empty. If the interior of either one of them is completely included in the interior of the other, inclusion has the value *total*. If the interiors only partly overlap each other, inclusion has the value *partial*. If they do not overlap at all, the value for inclusion is *none* (after Bédard, 1990).

***adjacency*** involves two objects sharing a subset of points, but with one main condition: if their interiors intersect then only the boundaries are considered. If these conditions are met, then adjacency is either partial or total.

**Topology**

| | |
|---|---|
| **subclass:** | **TopologyExtended** |
| attributes: | [dimensionalityOfIntersection] [encircles] [encircledBy] [sur rounds] [surroundedBy] |
| domains: | Integer Boolean Boolean Boolean Boolean |

Multi-dimensional topologic information is described through the *dimensionality of the intersection of two objects*, as well as four additional comparisons: *encircles*, *encircledBy*, *surrounds*, and *surroundedBy*.

***encircles:*** object A encircles object B if the (outer) boundary of A, or A itself, can serve as the boundary of an object A' which contains object B. That is, A' encircles B, if A'   B.

***encircledBy:*** A is encircledBy B if B encircles A.

***surrounds:*** object A surrounds object B if a subset of the boundary of A, or A itself: (i), does not intersect B, and (ii), serves as the entire boundary of a third object which encircles B.

***surroundedBy:*** A is surroundedBy B if B surrounds A.

***DimensionalityOfIntersection*** attribute, equal to 0, 1, 2, or 3, refers to whether the intersection of objects A and B is a member of PointClass (0), LineClass (1), AreaClass (2), or VolumeClass (3), respectively. If the intersection results in objects of different dimensionality, the largest number is provided.

The Topology defined in SAIF accomodates a variety of relationships which may be of interest which do not fit into a simpler approach without creating the need for numerous descriptors. A semantically complicated list would be required and it introduces the problem of appropriate names.

The SAIF Data Model provides structures for describing the temporal topology that exists between events, because temporal information is not always expressed in metric units (i.e the Renaissance happened ***after*** the Dark Ages and ***before*** the Industrial Age).

**Relationship**

| | |
|---|---|
| **subclass:** | **TemporalTopology** |
| attributes: | name |
| domains: | TemporalChoice |

If A occurred *before* B, then A is the primary object, B is the related object and *before* comes from TemporalChoice.

The choices for describing the Temporal relationship come from an enumeration called TemporalChoice. This enumeration gives all the valid means for comparing the relationship between events. Note that these relationships are similar to those used for currency control in computer science applications.

**Enumeration**

| | |
|---|---|
| **subclass:** | **TemporalChoice** |
| values: | equals before after meets metBy overlaps overlappedBy during contains starts startedBy finishes finishedBy |

This enumeration provides the choices for a comparison of the occurrence or observation of one object and that of another. These choices *(Allen, 1984)* all specify binary temporal relationships between a primary object, A, and a secondary object, B. Note that in the descriptions which follow, time is treated as a duration (from the domain Duration with a startTime and endTime). If time is defined only as part of a coordinate (e.g. from CoordXYT) or by the attribute timeStamp in GeometryExtended, only the first three values should be used (ie. if the notion of time is that of a single valued instant or snapshot, then valid choices are restricted to *equals*, *before*, and *after*.)

***equals:*** The startTimes of objects A and B are the same and the endTimes of the objects are also the same. (If time is considered as a single value and not a duration, then the time of A and the time of B have the same value.)

***before:*** The endTime of A precedes (i.e., is less than) the startTime of B. (If time is considered as a single value and not a duration, then the time of A precedes, or is less than, the time of B.)

***after:*** The endTime of B precedes the startTime of A. (If time is considered as a single value and not a duration, then the time of B precedes the time of A. ***meets:*** The endTime of A equals the startTime of B.

***metBy:*** The endTime of B equals the startTime of A.

***overlaps:*** The startTime of A precedes the Start-Timeof B; the startTime of B precedes the endTime of A; and the endTime of A precedes the endTime of B.

***overlappedBy:*** The startTime of B precedes the startTime of A; the startTime of A precedes the endTime of B; and the endTime of B precedes the endTime of A.

***during:*** The startTime of B precedes that of A, and the endTime of A precedes that of B.

***contains:*** The startTime of A precedes that of B, and the endTime of B precedes that of A.

***starts:*** The startTimes of A and B are equal, and the endTime of A precedes that of B.

***startedBy:*** The startTimes of A and B are equal, and the endTime of B precedes that of A.

***finishes:*** The endTimes of A and B are equal, and the startTime of B precedes that of A.

***finishedBy:*** The endTimes of A and B are equal, and the startTime of A precedes that of B.

These choices provide 13 different unambiguous temporal relationships between two objects.

Another temporal relationship deals with events that do not overlap but have a sense of "*neighbourhood*", in that there is an interval of time that passed between events. A TemporalOffset relationship exists which allows the description of an instant in relation to the start of an event or the end of an event.

**Relationship**

| | |
|---|---|
| **subclass:** | **TemporalOffset** |
| attributes: | instanceA instanceB from offset |
| domains: | TemporalInstance<br>TemporalInstance<br>BeforeOrAfter GeneralTime |

For A and B, it is specified whether the point of time in question is a startTime, an endTime or a single Instance ( the start and endTimes are assumed to be equivalent). The time instance of A may occur before or after that of B, by a given amount, the offset.

### Metadata and data quality

Information being used for Resource Management is, in most cases, imperfect. Forestry Inventory, Fisheries Management, Geological Exploration, Habitat Analysis, Wilderness Management and Environmental Impact Analysis all rely on interpolation and extrapolation of data. Many of the phenomena being studied are not completely accessible to the researcher, so that the size and orientation of an object or duration of an event cannot be assumed. Furthermore, the continuity and orientation of the boundaries of the geographic object cannot always be discretely quantified. The boundaries are modelled quantities requiring that a statistical error be identified and associated with the geographic object. Issues related to "Vague Data" include:

- recognition that in many cases data are discrete samples of continuous surfaces and are therefore, spatially and temporally dependent,
- data comes from a variety of sources resulting in varying degrees of reliability. Indicators of quality must be kept with the data,
- sparsity of data may result in the combination of information from a range of time and space. Data structures must exist to adequately reflect this fact in the synthesized data. This may require that a "history" of data manipulation take place as well as a quantification of associated error,
- the conventional Database Management Paradigm based on Relational Set Theory is not equipped to deal with *Fuzzy Data* (Wang, 1990),

The SAIF Data Model is structured to facilitate the direct association of uncertainty with any type of Geographic object. Much of the ability to express uncertainty relies on the structure of the Coordinate class structure (Figure 3). Other methods are available in SAIF for qualifying or quantifying uncertainty and data quality. Much of this information is contained in the Metadata component of a Geographic Object. An example of part of the SAIF data structure for georeferencing and storing data quality is given below.

**Tuple**

| | |
|---|---|
| **subclass:** | **MetaData** |
| attributes: | [georeferencing] [content]<br>[source] [history] [quality]<br>[addedClasses] [update]<br>[userMetaData] |
| domains: | Georeferencing Content Source<br>History Quality<br>List(TypeDefinition)<br>List(UpdateOperation) Tuple |

Metadata contains eight components which together provide information about one or more geographic objects. ***Georeferencing*** is the attribute which provides foundation to give meaning to the positional information. The ***Quality*** component addresses several data quality concepts.

The **Georeferencing** tuple provides key information regarding cartographic location and accuracy of the geographic information. The structure of the Georeferencing Subclass has the following components;

*coordSystem:* concerns the way horizontal and vertical coordinates are defined. Attributes include: Projection, Origin and Units (ie. Cartesian, Geocentric and metres.)

*geodeticReference:* provides ellipsoid and vertical datum information.

***hydrographic:*** tidal and sounding information. Control supplies vertical and horizontal control information.

***transform:*** Supplies transformation criteria for the generation of corrected coordinates when data has been transformed; ( e.g. if the coordinates have been scaled to values between 0 and 1, or if the coordinates are shifted relative to some starting point).

***location:*** a bounding box in 2, 3, or 4 dimensions may be specified, as well as the tiles (mapsheets) which include the object.

***remarks:*** comments can also be provided.

The **Quality** component of metadata contains a series of attributes which characterize different aspects of data quality.

***structuring:*** this specifies a series of attributes related to the notion of planar enforcement in two or three dimensional space. These include shared points and boundaries, space filling (including overlapping), and void space within the spatial extent of the data.

***positional:*** specifies absolute and relative accuracies pertinent to coordinate data.

***attribute:*** specifies statistical measures associated with the thematic content of the data, i.e., content as opposed to positional accuracy.

***incomplete:*** concerns size criteria for classes of objects, instances smaller than the criteria are either not present in the data or are represented as objects of lower dimensionality (such as a river being represented by an instance from LineClass, instead of from AreaClass or VolumeClass).

***clipping:*** indicates if any objects have been truncated in order to meet the data transfer request.

***integrity:*** provides the results from a series of data quality tests.

## Conclusion

SAIF provides a strong foundation upon which to structure information for use in multi-dimensional or multi-temporal analysis. The model is based on the object oriented paradigm in a way that we believe is mappable onto an extended relational model. We have endeavoured to balance simplicity of design with utility of product. To maintain this simplicity, we have limited the model to single vs. multiple inheritance among objects, while allowing user extensibility of the schema within certain constraints and guidelines.

As providers of a data interchange standard, the Surveys and Resource Mapping Branch believes it is imperative that SAIF be able to accommodate the following:

- correct storage and manipulation of data that is referenced in 2,3 and 4 dimensions,
- provision of structures that can associate uncertainty with georeferenced information,
- retrieval and updating of temporal information,
- provide data structures appropriate for designing real-time systems (ie. navigation),
- use of artificial intelligence in planning and reasoning about temporal data,
- software Engineering in the design and specification of properties of Open Systems and distributed database technology,
- enterprise-wide data access,

These functions may be beyond the scope of requirements for a GIS but they are not beyond requirements of a corporate information management system. SAIF is designed to fulfil the need for a provincial and national standard for the transfer of geomatics information.

## Bibliography

Allen, J.A., (1983), Maintaining Knowledge About Temporal Intervals, Communications of the ACM, Vol. 26, No. 11, November, pg 832 - 843.

Bédard,Y., (1990), On Spatial Modelling, Laboratoire de SIRS, Centre de Géomatique, Départment des sciences géodésiques et de télédétection, Université Laval, Sainte-Foy, 26 pages.

Bell, S. M. B., Diaz, B.M., Holroyd, F.C., (1988), "Capturing Image Syntax Using Tesseral Addressing and Arithmetic", Digital Image Processing in Remote Sensing, edited by Muller, J-P, Taylor and Francis, London

Coad, P., and Yourdon, E.,(1989), "Object Oriented Analysis", *Systems and Software Requirements*, Pages 272 - 288.

Coad, P., and Yourdon, E., *Object Oriented Analysis*, Prentice - Hall, 1990

Forrester, W.D., (1983), Canadian Tidal Manual, Dept. of Fisheries and Oceans, Canadian Hydrographic Service, Ottawa, Ont., 138 pgs.

Guide to GPS Positioning, (1986), edited D. Wells et al, Canadian GPS Associates, Fredricton, NewBrunswick, University of New Brunswick Graphic Services.

Maiocchi, R., Pernici, B., (1991), Temporal Data Management Systems: A Comparative View, IEEE Transactions on Knowledge and Data Engineering, Vol. 3, No. 4, December, pg 504 - 523

Raper, J., (1989), *Three-Dimensional Applications in Geographic Information Systems*, Taylor and Francis, Ris, F., (1991), "Toward An 'Scientific GIS'", *What Is Scientific GIS?*, IBM Conference Proceedings, September 19, 1991., Boulder, Colorado.

Samet H., (1990), Design and Analysis of Spatial Data Structures, (1989), Addison - Wesley, New York, 489 pages

Samet, H., (1990), Application of Spatial Data Structures, Computer Graphics, Image Processing and GIS, Addison - Wesley, New York, 505 pages 188 pages

Wang, F., Hall, B.G., Subaryono, (1990), Fuzzy Information Representation and Processing In Conventional GIS Software: Database Design and Applicaton, *International Journal Of GIS*, Vol. 4, No. 3, Pg. 261 -283.

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# NAD27 to NAD83 conversion in British Columbia: national transformation, version 1.1

**Amin M. Kassam, P.Eng. A.L.S.**
**Kent W. Pointon, P. Eng.**
Head, Geodetic Control Unit
Surveys and Resource Mapping Branch
Ministry of Environment, Lands and Parks
4 th Floor, 1802 Douglas Street
Victoria, B.C. V8V 1X4

## Abstract

*Over the past several decades voluminous spatial and associated attribute data have been gathered by industry and government for a multitude of purposes. These data have usually been georeferenced to the provincial geodetic control system, either directly or, as is the case more often, indirectly through various base mapping. The Geodetic Reference System provides the coordinate framework for this georeferencing. In its horizontal component, the Geodetic Reference System has largely been based on the North American Datum of 1927 (NAD27). In May, 1990, the federal and British Columbia government announced the adoption of the new North American Datum of 1983 (NAD83) as the basis for the Geodetic Reference System in Canada and British Columbia, respectively.*

*The adoption and implementation of NAD83 is now well underway in British Columbia. Software has been developed to help users transform point coordinates in either geographic or Universal Transverse Mercator coordinates from NAD27 to NAD83. This paper briefly backgrounds the provincial geodetic control system; next, it describes at length the methodology used in devising this transformation process; and finally, it provides advice on the transformation's applicability for GIS/mapping data conversion. The transformation analysis shows the mean difference between adjusted and transformed coordinates of geodetic control monuments is 0.51 metres with a 90-th percentile difference of 1.14 metres. However, the analysis also shows that differences vary significantly throughout the province. Areas of the province where special caution must be exercised when using the transformation are identified. In conclusion, the paper shows that, generally speaking, the transformation is adequate for scales of mapping 1:10 000 and smaller.*

## Introduction

1990 marked a milestone in Canada and British Columbia with the publication of the higher order geodetic control networks coordinates on NAD83 for the very first time. Efforts are now underway, both in government and in private industry, to integrate more data onto thenew NAD83 foundation. These efforts are guided by *business decisions that evaluate the benefits of NAD83 versus the costs of remaining on NAD27 and the costs of conversion.* The National Transformation methodology presented in this document is therefore offered as a key tool for converting existing NAD27 data holdings of an organization.

The NAD83 system is replacing the NAD27 system which has been in use for the better part of this century. The large distortions in the NAD27 system are being removed since the NAD83 system is internally consistent - as a result of a simultaneous readjustment of all the horizontal geodetic control networks - as well as being accurate due to external satellite based data. To this end, NAD83 is consistent with the satellite survey techniques now being implemented (for example, the Global Positioning System or GPS).

As of October, 1990, NAD83 values for all but the urban geodetic control points in the province have been published and made available to the public. Currently, the integration of the urban geodetic networks into the NAD83 system is underway and official values will be available by March 31, 1993. Much background information on NAD83 and the communication efforts with the user community may be found in such mediums as the proceedings of the 1988 and 1990 Canadian Institute of Surveying and Mapping (CISM) NAD83 seminars. For a description of the NAD83 technical implementation strategy in the Province, see Kassam, et al, 1990.

### 1.1 Why NAD83 ?

NAD27, established through surveys across Canada with considerable care and within the limits of accuracy that the technology of the day had to offer, adequately served its intended purpose up to the early 1960's. The 1927 geodetic reference system cannot meet modern day demands of current and future position-related activities. The new NAD83 geodetic reference system is intended to meet these needs. Changes resulting from NAD 83 are therefore critical to future economics in the production and use of land-related information ( see for example, O'Brien, 1985).

It is not within the scope of this document to discuss the merits of NAD83. Several references are cited for the interested user. Nonetheless, it is worthwhile noting that within the B.C. government alone some $300 million are committed to programs relying on NAD83. Clearly, within this province, NAD27 simply will not meet the present and emerging demands for accurate and universally compatible spatially referenced information.

To recap, the following reasons are often cited as motivation to ensure conversion of all land-related data holdings (both public and private) to NAD83.

- Proliferation of local coordinate systems impede overlaying and edge-matching of information gathered from various sources (for example, utilities companies and municipalities). Universal compatibility is the foundation of true Geographic Information Systems (GIS). NAD83 provides for a homogeneous, global, survey control network on which all geo-referenced data can be made universally compatible.
- Distortions in the NAD27 datum based survey control coordinates result in added costs for surveys due to reliability concerns.
- Advent of Global Positioning System (GPS) technology has spurred myriad applications based on a niversal (global) coordinate system (that is compatible with NAD83).
- Future Active Control System (ACS) will be based on geocentric coordinate system consistent with the new NAD83 datum. Thus, highly accurate and reliable data collected tomorrow from cheap, hand-held, receivers by practically any citizen, will yield coordinates directly consistent with NAD83.
- The current Terrain Resource Information Management (TRIM) 1:20 000 mapping is already on NAD83. All future Federal Government mapping will also be undertaken on NAD83.
- The Canadian Government declared, on May 23, 1990, NAD83 as the official datum for Canada. The British Columbia Government followed immediately and announced adoption of NAD83 on May 24,1990. NAD83 is already adopted by the United States.
- Transformation of NAD83 compatible coordinates to old NAD27 datum based coordinates is an imprecise procedure due to the un-modelable distortions inherent in the NAD27 system.
- NAD83 conversion offers an opportunity to users to rationalize and computerize their current operations and clean-up their data resources.
- Low-cost user-friendly intelligent graphics work stations, in the future, will become common and users will demand from government agencies on-line access to consistent geographic data sets for their displays and decision-making.
- Costs associated with the maintenance of Integrated Survey Area (ISA) control points will be reduced due to more homogeneous survey control networks on NAD83 and the impact of GPS.
- Conversion software, from NAD27 coordinates to NAD83 coordinates, has been developed by the Federal Government and the Province so as to provide NAD27 users a means for data conversion.
- More control points will be available on NAD83 than on NAD27. New points will be computed on NAD83 only by both Federal and Provincial Governments because transformation to NAD27 (once surveyed under NAD83) is an inexact science. Old NAD27 coordinates will be phased out by the mid-1990's.

In order to understand the requirements for NAD83, we need to understand the uses for geodetic control. The following definition of geodetic control is set forth for these purposes. As well, the translation of the definition to the actual situation in British Columbia is presented next so as to offer contextual understanding of geodetic control usage in the province.

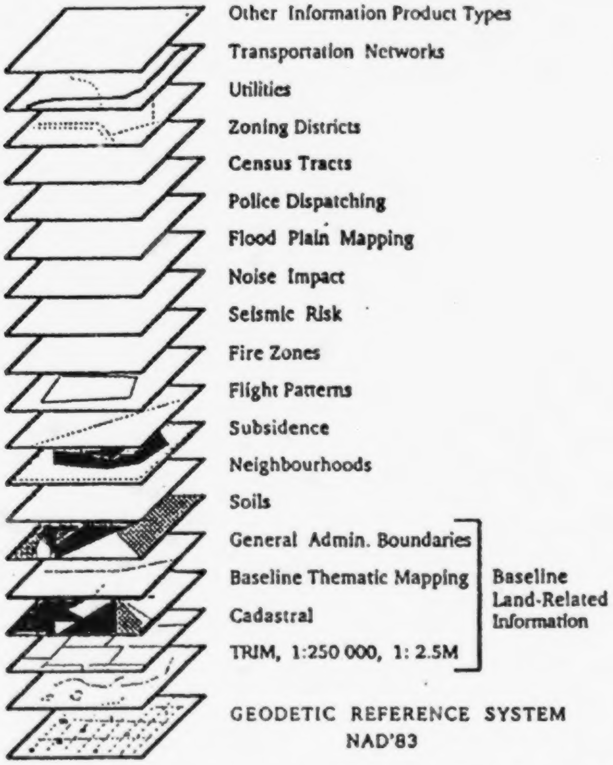

Other Information Product Types
Transportation Networks
Utilities
Zoning Districts
Census Tracts
Police Dispatching
Flood Plain Mapping
Noise Impact
Seismic Risk
Fire Zones
Flight Patterns
Subsidence
Neighbourhoods
Soils
General Admin. Boundaries
Baseline Thematic Mapping
Cadastral
TRIM, 1:250 000, 1: 2.5M
Baseline Land-Related Information
GEODETIC REFERENCE SYSTEM
NAD'83

**Figure 1:** *Secondary and tertiary products based on a Geodetic Reference System*

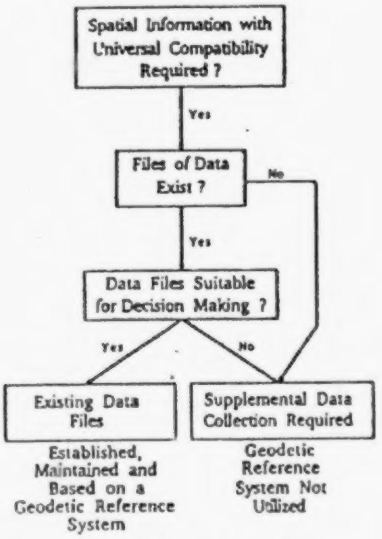

Spatial Information with Universal Compatibility Required ?
Yes
Files of Data Exist ?
No
Yes
Data Files Suitable for Decision Making ?
Yes
No
Existing Data Files
Supplemental Data Collection Required
Established, Maintained and Based on a Geodetic Reference System
Geodetic Reference System Not Utilized

**Figure 2:** *The land related decision process*

### 1.2 Geodetic control — definition

Geodetic Control, often referred to as Survey Control, can be compared to the steel skeleton which supports a building. As a steel skeleton supports a building, Geodetic Control, through its provision of an accurate coordinate system, supports all position-related activities. Position-related activities include, but are not limited to (Kassam, 1989):

- Development and operation of computerized land-related information systems
- Topographical and resource mapping
- Cadastral (real properties) mapping
- Integrated engineering projects (planning, reconnaissance and construction)
- Exploration for and recovery of natural resources
- Coordinating positions of surface / subsurface objects for future location and protection
- All forms of navigation (water, air, land)

***In the same way that the multi-storied building requires a substantial foundation and an accurately positioned steel framework, modern positional applications (e.g. Geographic Information Systems or GIS's) require a precise coordinate system framework*** (Figure 1). It is also obvious from the preceding that secondary and tertiary users of the geodetic reference system avoid costs of compatibility with other information products when attempting to aggregate or compare data and information. Figure 2 illustrates this point (University of Maine at Orono, 1984).

### 1.3 Geodetic control — application

The Ministry of Environment, Lands and Parks has the mandate to manage the geodetic reference for subsequent geographic positioning and navigation. Further, the Ministry is tasked to secure geographically referenced baseline data fundamental to the GIS requirements of British Columbians engaged in managing natural resources and municipal infrastructure, and to manage a corporate GIS framework within which GIS's can evolve within government ministries while maximizing opportunities for data sharing.

The geodetic reference management requirement within the Ministry also derives from legislation under the Land Survey Act and the Petroleum and Natural Gas Act. These Actsare fundamental to the cadastral fabric of the province and to exploration activities in the petroleum and natural gas industries, respectively.

In British Columbia, the Land Information Infrastructure (L.I.I.) is a concept whereby land-related activities can be supported through a (largely) computer-based medium. Integral to this medium is the need for universal (data) compatibility, entailing a standard base referencing system with standard technology (i.e. compatibility, not necessarily same hardware / software ! ) and a standard access system for the exchange of land-related information. Given such a medium, baseline land-related information - that is the information system forming the base land-related data describing the B.C. water and land masses - may be linked to subject data bases containing location based information to facilitate integrated business decisions using GIS technology (Figure 3; Kenk, et al, 1990).

The implementation of this L.I.I. in the province is being directed by the Land Information Strategic Committee and it is envisaged that the benefits of this program will extend to local governments and the private sector in the form of more timely and accurate data exchange as well as the provision of better land information (Duffy, 1990).

Survey Control provides the basic geodetic reference system necessary for an accurate and consistent framework in the L.I.I. NAD83 is the standard for modern georeferencing needs within this L.I.I.

## NAD27 to NAD83 conversion: National transformation approach

The positions of monuments estimated on NAD83 are far more internally consistent than the corresponding NAD27 positions. Much of the high level of distortion present in the NAD27 adjustment system has been removed under NAD83 and the NAD83 datum is consistent with the new satellite based survey techniques. The level of distortion in NAD27 has been discussed in several papers, for example, see: Pinch (1990), Wilkinson (1988).

The Surveys and Resource Mapping Branch recognized the value of the NAD83 readjustment early and since 1986 has based the TRIM 1:20 000 mapping program on preliminary coordinates from the continental adjustment. All new surveys performed for the Surveys and Resource Mapping Branch of The Ministry of Lands and Parks and the other Ministries of the Province of British Columbia will generally be referenced to the new NAD83 adjustment system. *The current control monument data holdings on NAD27 will be supported for a period of up to five years from first publication of NAD83 coordinates, but no new surveys will be computed on the old NAD27 datum.*

The NAD27 adjustment system has been in use for almost 60 years. Most of the existing positional data is referenced to the NAD27 adjustment system. Converting this stored positional data to the new NAD83 datum will be a costly and time consuming process. To achieve the utmost in accuracy and reliability, a rigorous adjustment of all the survey observations would have to be carried out. This adjustment method is what the Surveys and Resource Mapping Branch is using to determine the NAD83 coordinates of the remaining geodetic control networks, largely comprising of the ISA networks. These

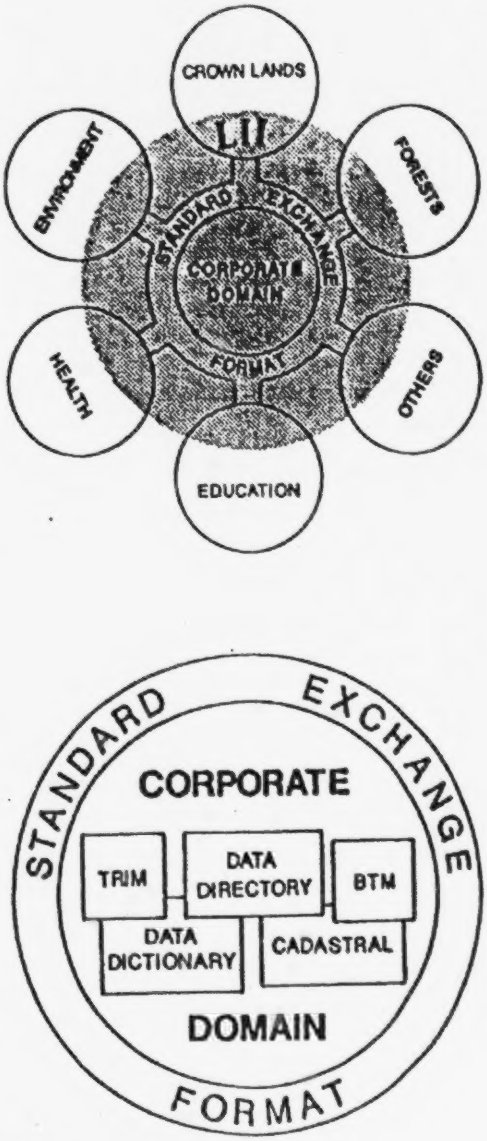

CROWN LANDS
LII
ENVIRONMENT
FORESTS
STANDARD EXCHANGE FORMAT
CORPORATE DOMAIN
HEALTH
OTHERS
EDUCATION
STANDARD EXCHANGE FORMAT
CORPORATE
TRIM
DATA DIRECTORY
BTM
DATA DICTIONARY
CADASTRAL
DOMAIN

**Figure 3:** *The Land Information Infrastructure (L.I.I.). The L.I.I.* facilitates the sharing *and exchange of land related data. The Corporate Domain at the hub of the L.I.I. consists of baseline data that can be utilized by a number of user agencies. The Standard Exchange Format will provide the means through which data is exchanged between ministries.*

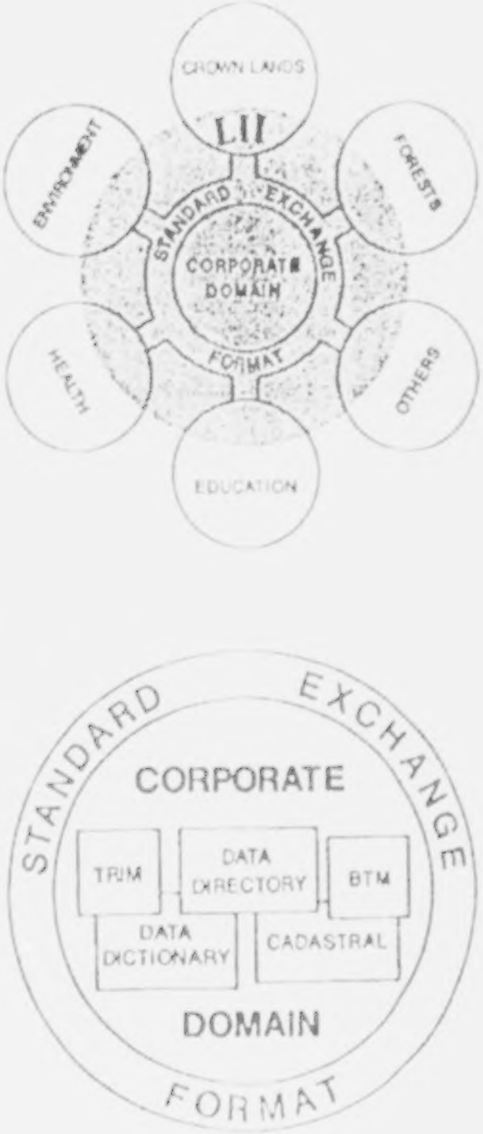

CROWN LANDS
FORESTS
OTHERS
EDUCATION
HEALTH
ENVIRONMENT
LII
STANDARD EXCHANGE FORMAT
CORPORATE DOMAIN
STANDARD EXCHANGE FORMAT
CORPORATE
TRIM
DATA DIRECTORY
BTM
DATA DICTIONARY
CADASTRAL
DOMAIN

**Figure 5.** *The Land Information Infrastructure (L.I.I.). The L.I.I. facilitates the management of land and land related data. The Corporate Domain at the hub of the L.I.I. consists of [illegible] accessed by a number of user agencies. The Standard Exchange Format will provide the [illegible] between [illegible] agencies.*

networks will be integrated onto NAD83 through a province-wide readjustment process by April 1, 1993.

*Performing a rigorous adjustment is simply infeasible for the majority of thematic, cartographic, and lower accuracy network data.* In response to the need for converting this data, a public domain software package has been prepared by the Canada Centre for Surveying, in conjunction with the Canadian Control Survey Committee (CCSC), to provide an approximate transformation. This transformation package (called the National Transformation) has been adopted as a national standard by the CCSC and will be widely distributed. Utilizing this package will ensure that all your transformed coordinates will be compatible with anyone else using the package Other transformation methods are available (or may be derived), but the compatibility with the National Transformation will be lost. However, local transformation accuracy may be improved. Agencies contemplating transformation programs should carry out a detailed study of the region of interest.

The National Transformation is based on a grid of shifts that, when applied to NAD27 coordinates, will give an approximate NAD83 coordinate. Bilinear interpolation routines are provided to determine a unique shift at any point within the grid. This grid of datum shifts (at an interval of five arc minutes) and the interpolation routines essentially define a transformation surface. This surface exhibits several topological properties:

1. It is continuous and the grid boundaries are collinear;
2. Transformations of continuous features will remain continuous both within the grid and across grid boundaries; and
3. Linear features will be distorted, i.e. conformality is not preserved (Junkins, 1989).

Details of the design and implementation of the national transformation software can be found in Junkins (1990), Farley and Junkins (1990) and Junkins (1988).

The following sections quantify the level of error that this transformation imparts and provide specific guidance to provincial users of the National Transformation for their particular applications. For a description of the software and the associated Users Guide, refer to the Canadian Institute of Surveying and Mapping (CISM) publication 'Moving to NAD83 - the new address for georeferenced data in Canada'(1990).

## National transformation surface definition issues

Coordinates for approximately 14 000 monuments in British Columbia were determined in the continental readjustment. The Surveys and Resource Mapping Branch have coordinates in the NAD27 adjustment system for most of these monuments. The exception being monuments established for additional control of the TRIM mapsheets (approximately 700 monuments) which have only NAD83 coordinates. Of the remaining monuments, approximately 1 450 were originally classified as Fourth Order Horizontal. These monuments were excluded from the definition of the transformation surface in view of their limited accuracy. In addition, 2 500 First to Third Order monuments were excluded due to large differences between transformed and adjusted coordinates or incompatibility with surrounding networks. Various criteria were used to determine which monuments were to be excluded, large differences between the transformed coordinates and the adjusted coordinates being the primary, but not only criteria.

Figures 4 and 5 show the final transformation surface definition as broken down into its latitude and longitude components. Thus, each contour line links together points of equal latitude (Figure 4) or longitude (Figure 5) shift from NAD27 to NAD83. The contour intervals are 0.2 arc seconds, representing roughly 5 metres. These figures give a good broad brush view of the surface characteristics (slope trends and regional differences) and is not intended for scaling off particular shifts from NAD27 to NAD83.

The final accuracy of this type of transformation is subject to the homogeneity of the two data sets (coordinates of monuments in both the NAD27 and NAD83 systems). Figure 6 depicts a transformation surface derived from a consistent set of monument coordinates. As indicated here, a coordinate shift from NAD27 to NAD83 at any one point between grid points will still be slightly in error due to the difference between the curved surface and the linear interpolation but the magnitude will not be excessive (the actual difference being proportional to the spacing between the control monuments).

Any distortions or local inconsistencies in either system will distort the transformation surface. The coordinates of monuments on NAD27 are notorious for local distortions, inaccuracies and unintegrated networks. Figure 7 shows how the transformation surface will be defined in such a case. Here, the surface is placed at a best fit mean, but it does not provide a reasonable transformation for either networks' coordinates.

Since all survey measurements are subject to errors and all adjustments are minimum variance estimates of the true position of the monuments, the scenario shown in figure 7 is the rule rather than the exception. Before releasing version 1.1 of the National

Transformation, a significant amount of work with the test version 1.0 was put into reducing the magnitude of the errors at monuments of more recent or more often used networks. Errors at monuments and networks which show local distortions were allowed to increase to

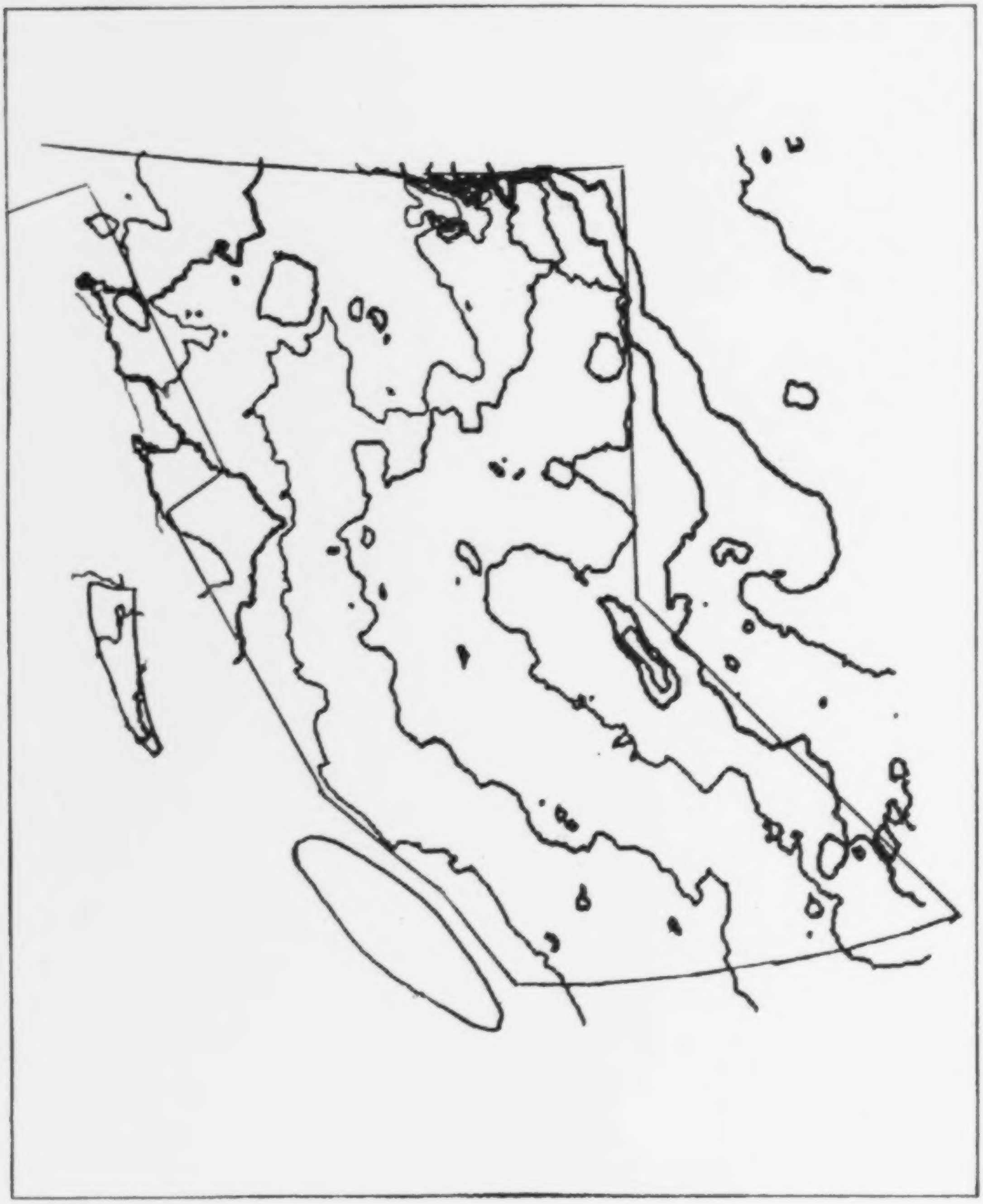


**Figure 4:** *Transformation surface latitude contours (0.2" intervals)*

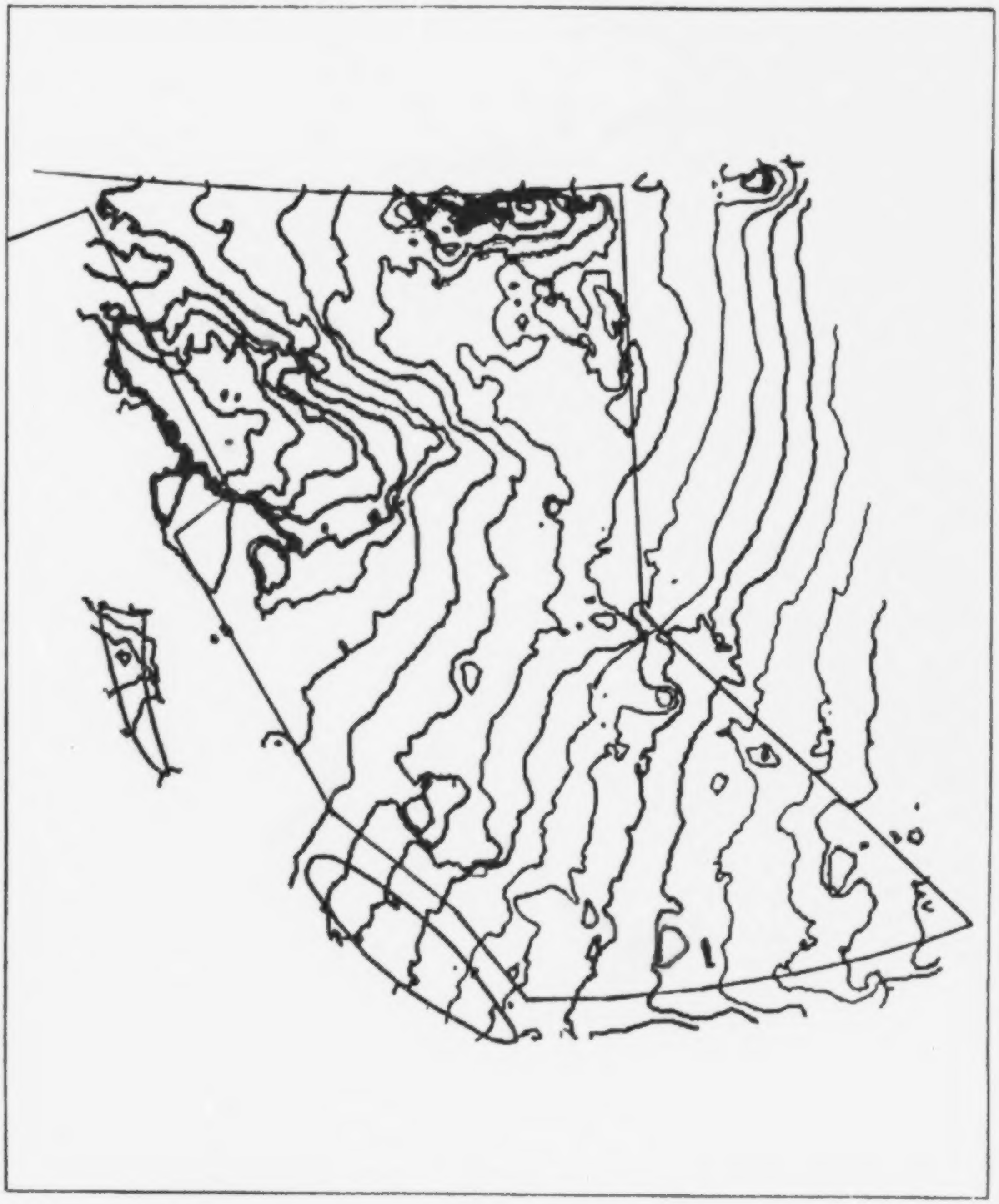


**Figure 5:** *Transformation surface longitude contours (0.2°intervals)*

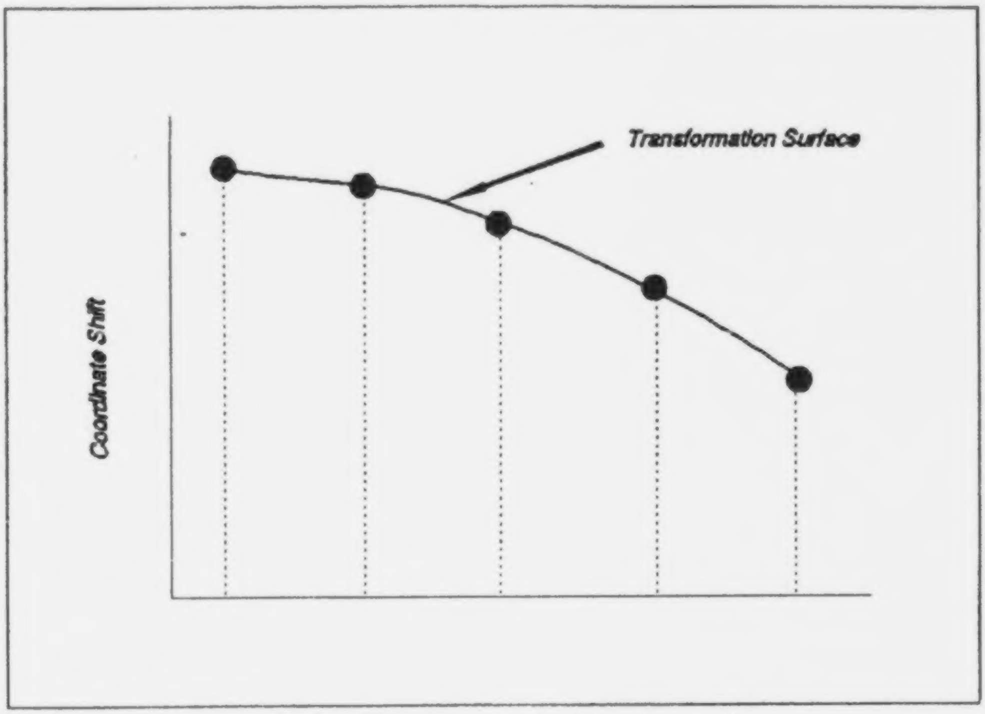

Transformation Surface
Coordinate Shift

**Figure 6:** *Transformation surface defined by consistent coordinates*

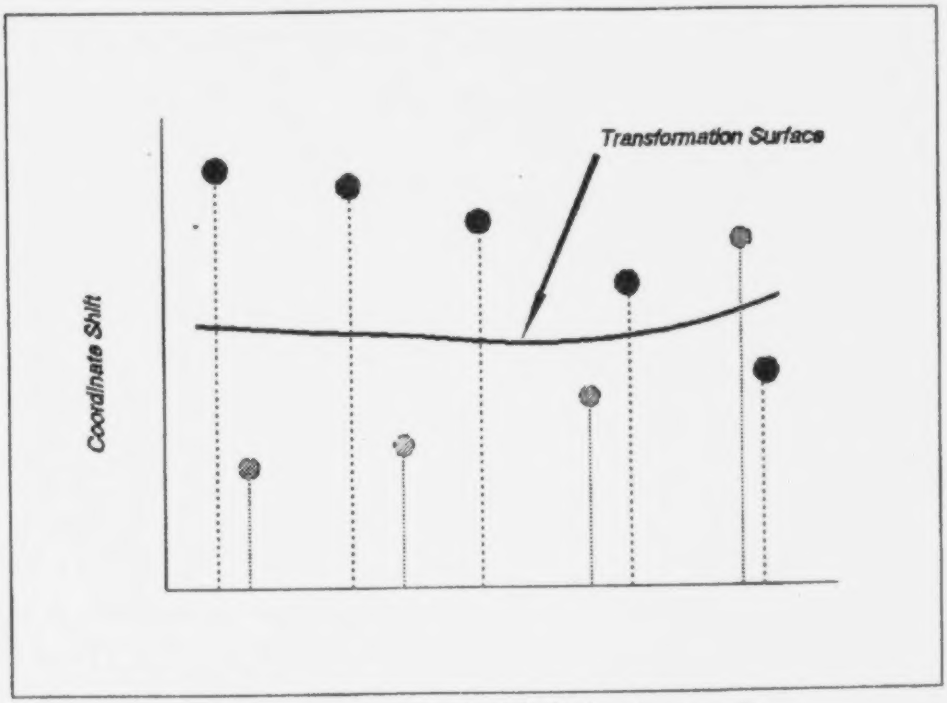

Transformation Surface
Coordinate Shift

**Figure 7:** *Transformation surface defined by inconsistent networks*

provide a better fit to the more consistent networks. This is accomplished by excluding the monuments which show local distortions from the definition of the transformation surface. In the following sections 3.1 to 3.5 the criteria used in excluding the monuments is discussed. Details of particular areas of concern will be discussed later. Section 3.6 discusses the issues of country-wide definition of the transformation surface, whereas Section 3.7 explains the definition issues in the municipal Integrated Survey Areas.

### 3.1 Large differences between transformed and adjusted coordinates at single points

Differences between the adjusted and transformed coordinates of a single monument of up to 126 metres were seen in the initial definition of the surface. Residuals of this magnitude primarily occurred at weakly determined monuments. They may have been established by side shots or by intersection of observed lines where the stations were not occupied (i.e. towers). All identified stations of this type were removed from the final definition of the transformation surface.

### 3.2 Local distortions and incompatible networks

In several areas of the province, two or more networks of different vintage can be found in the same area. For various reasons, these networks were never tied together. Figure 8 depicts this scenario. Depending on how the initial adjustments were done and the consistency of the coordinates of the monuments used as fixed control in the two networks, inconsistencies (localized datums and significant distortions between monuments of the two networks) can occur. This is exactly the problem shown previously in Figure 7.

One of the side benefits of the NAD83 continental readjustment and this transformation surface definition exercise has been the identification of several of these inconsistent areas. The problem primarily exists where there are networks established by provincial and federal agencies in the same area. As a general rule, monuments from the federal networks were excluded from the definition of the transformation surface, thus providing a better fit to the higher density provincial networks. All federal Mapping and Charting Establishment (MCE) monuments, National Defence (NATDEF) and Dominion Topographic Survey (DTS) monuments were excluded (even though they may have had a higher horizontal order designation) since the provincial networks generally exhibited better overall fit in the same areas.

The problem of incompatible networks was also evident with several federal First Order networks which were put in after the provincial networks. Federal monuments were put on the same hill tops but not tied to the provincial networks. Even though the distance between the monuments is less than 100 metres, incompatibilities of up to 30 metres in the NAD27 coordinates were identified. Problem networks of this type occurred in Quadsheets 93, 94 and 104. Networks in the area of Williston Lake in and near NTS Map sheet

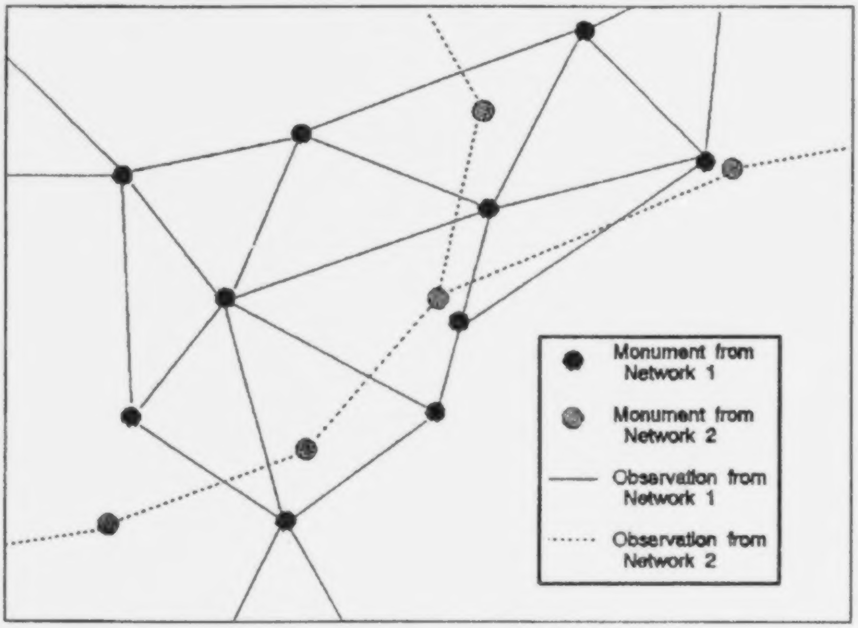

● Monument from Network 1
● Monument from Network 2
—— Observation from Network 1
······ Observation from Network 2

**Figure 8:** *Overlapping networks*

94/C with Unique identifiers 757nnn and the region and the federal networks of the Unique Identifier series 737nnn and 747nnn in NTS Mapsheet 94/D are significant problem areas of this type.

If a user has a network based originally on the coordinates of these monuments, significant errors in the transformed coordinates can result. This area of the province has shown particularly poor integration of the NAD27 networks. Users of the transformation software should exercise due caution in these regions.

### 3.3 Boundary monuments

Coordinates for the boundary monuments in several areas of the province were found to be inconsistent with the surrounding control. Therefore, all boundary monuments were excluded from the definition of the transformation surface. The exclusion included all International Boundary Commission (IBC) monuments, all BC-Alberta, BC-Yukon and BC-Northwest Territories monuments. The BC-Alberta and BC-Northwest Territories had coordinate differences which approached twenty metres and more in many cases. The BC-Yukon and IBC monuments were, in general, quite good, with most coordinate differences less than two metres. However, large differences do occur for these monuments.

### 3.4 Weakly determined networks

Several local networks which are weakly tied to the surrounding control have been identified, for example at Kinbasket Lake, Taweel Lake and Horsefly Lakes. Coordinates for monuments from these networks are weak in both NAD27 and NAD83, thus they were excluded from the definition of the transformation surface.

### 3.5 Unmarked monuments

Monuments without permanent marks usually carry a Unique Identifier of the form yyHNnnn, where yy is the year of installation, H is an identifier for British Columbia, N means not marked, and nnn is an identification number. Entire networks of these monuments were excluded from the definition of the transformation surface in an attempt to get a better fit of the surface to coordinates of the other networks in the area.

### 3.6 Country-wide compatibility

The transformation surface has been defined nationally through member agencies of the Canadian Control Survey Committee (CCSC). Some 80 000 control monuments extending across the country were used in this process. Thus, complete integrity and compatibility across provincial and territorial boundaries has been achieved, the exception being in the Offshore areas of Canada (currently being addressed). However, each jurisdiction (provinces, federal government) is responsible for officially endorsing the use of the National Transformation within its own borders. *Users are therefore advised to consult with these jurisdictions before attempting to use the software outside of British Columbia.*

### 3.7 ISA monuments

The current published NAD83 control monument data set does not include the Integrated Survey Areas (ISAs) monuments. Thus, some 25 000 monuments in about 50 urban centres in the province have yet to be readjusted onto NAD83. These are currently being readjusted onto NAD83 and publication of coordinates is targeted for April 1, 1993. In the meanwhile, only a handful of control monuments define the transformation surface of the National Transformation in each of these urban areas.

It is expected that the final NAD83 coordinates within the ISAs will be within 0.5 m. of the current values for the limited data sets currently available. Therefore, one should not use the National Transformation in urban areas for those applications requiring 0.5 m. accuracies. *Large scale mapping conversions, such as 1:500 and 1:1 000 scales, will most definitely not be supported by the current version 1.1 of the National Transformation.*

Refer also to section 6.0 for a discussion on transforming in urban areas.

## Transformation surface error level statistics

A data set of 9272 monuments with horizontal order classifications of First to Third on NAD27 was used to determine the difference between the transformed and adjusted NAD83 coordinates. This data set does not include monuments excluded from the definition of the transformation surface. For this 9272 monument data set, the following statistics apply:

| | |
|---|---|
| Mean Coordinate Difference (m): | 0.51 |
| 90-th Percentile (m): | 1.14 |
| Maximum Difference (m): | 20.92 |
| Standard Deviation (m): | 0.72 |

A mean coordinate difference of one-half a metre should provide reasonable transformation accuracy for many applications such as GIS and cartography. Users requiring higher accuracy for applications such as geodetic positioning should look into readjusting the original observations or re-acquiring observational data if this is justifiable. However, the National Transformation does produce transformed coordinates that are accurate to within a few centimetres with respect to the surrounding control in several regions of the province. The converse may also be true, therefore, users should complete their own analysis in the region of interest before any major transformation project is undertaken. The quality of the transformations will be directly related to the internal and external consistency of the monument coordinates, both in NAD27 and NAD83, used to derive the transformation surface. *The National Transformation cannot, nor will any procedure, improve the quality of inferior data (Dewhurst, 1990).*

### 4.1 Statistical breakdown by NTS 1:25 000 mapsheets

Figure 9 shows the Mean coordinate differences broken down by NTS Quadsheet. Figure 10 lists the 90-th percentile of the magnitude of the coordinate differences, broken down by NTS 1:250 000 mapsheet. Table 1 provides a more extensive list of statistics, this time broken down by NTS 1:250 000 mapsheet. The 90-th percentile figures shown in Table 1, indicate that 90% of the monuments in that region have a difference between the transformed and adjusted NAD83 coordinates less than or equal to the 90-th percentile. For example, in mapsheet 82/E, the coordinate difference was less than or equal to 0.93 metres for 90% of the monuments in this mapsheet. The 90-th percentile and the mean coordinate differences do not include the monuments excluded from the definition of the transformation surface.

### 4.2 Excluded monuments statistics

The list of coordinates excluded from the definition of the transformation surface and the differences between the transformed and adjusted coordinates is available from Surveys and Resource Mapping Branch as part of the National Transformation package. This list and accompanying statistics should aid in determining what can be achieved from the transformation. In general, networks in the region of the monuments which exhibit large coordinate differences may prove problematic, especially if the network is tied directly to one of these monuments.

### 4.3 Local versus large regions relative accuracies

Users should also note that the error level tends to be relatively constant over a limited spacial extent. The relative accuracy within a network covering a small region will not be affected to the same extent as a network covering a major portion of the province. The transformed NAD83 coordinates of the monuments may only be accurate to within a half metre relative to the surrounding control, but the relative accuracy within the network will be preserved to a great extent. The

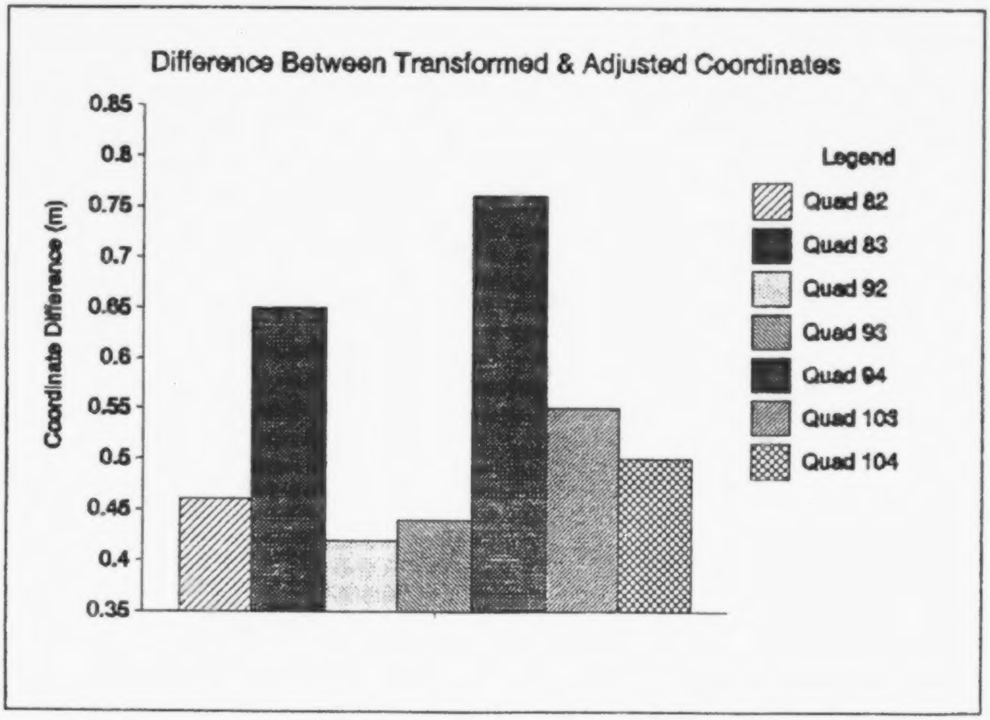


**Figure 9:** *Mean coordinate differences by quadsheet*

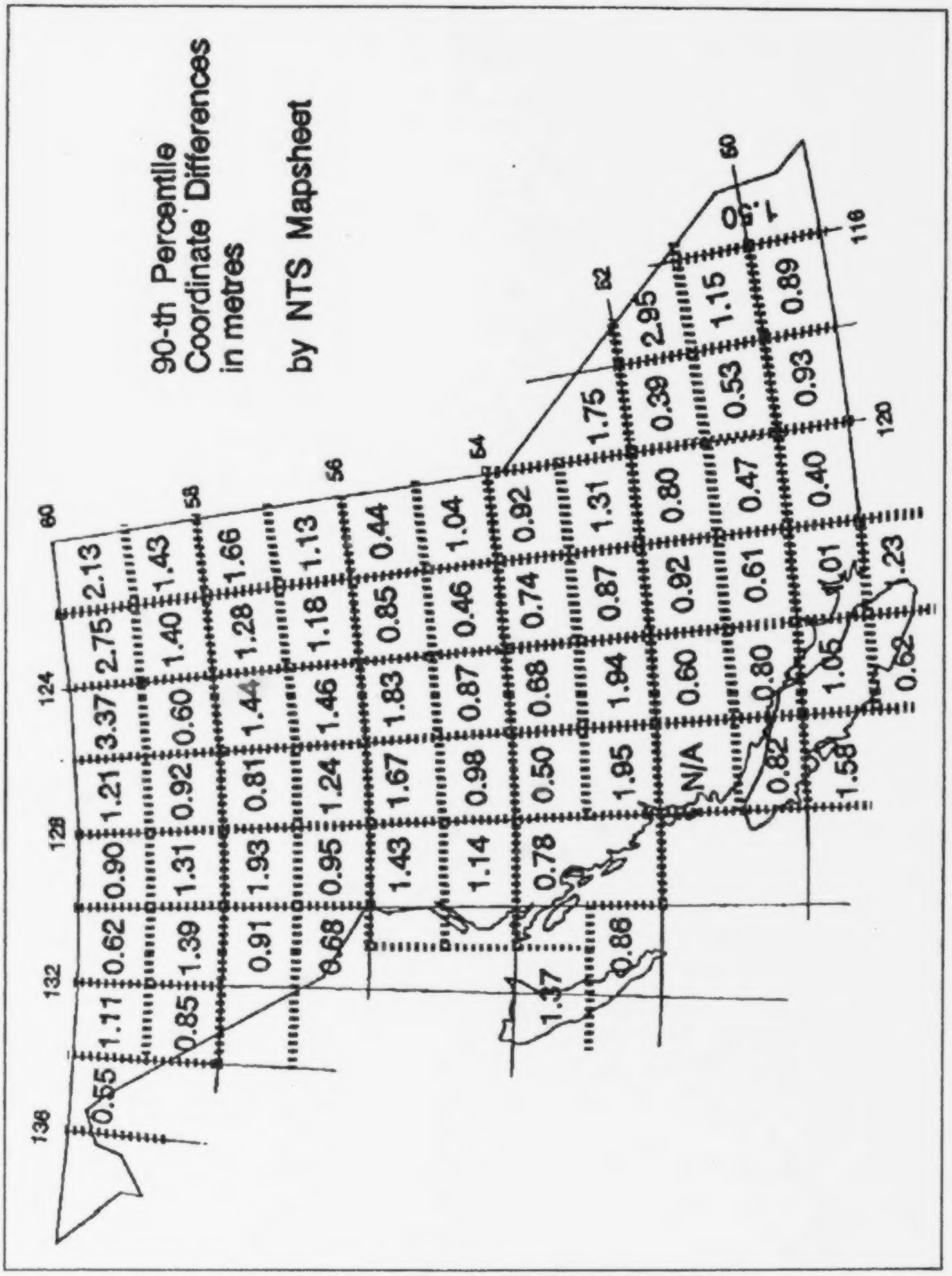

90-th Percentile
Coordinate Differences
in metres
by NTS Mapsheet
136
132
128
124
120
116
60
58
56
54
52
50
0.55
1.11
0.62
0.90
1.21
3.37
2.75
2.13
0.85
1.39
1.31
0.92
0.60
1.40
1.43
0.91
1.93
0.81
1.44
1.28
1.66
0.68
0.95
1.24
1.46
1.18
1.13
1.43
1.67
1.83
0.85
0.44
1.14
0.98
0.87
0.46
1.04
1.37
0.78
0.50
0.68
0.74
0.92
0.86
1.95
1.94
0.87
1.31
1.75
N/A
0.60
0.92
0.80
0.39
2.95
0.82
0.80
0.61
0.47
0.53
1.15
1.58
1.05
1.01
0.40
0.93
0.89
1.50
0.62
1.23

**Figure 10:** *90th percentile coordinate differences by NTS mapsheet*

**Table 1: Statistics Listed by Mapsheet for 9272 First to Third Order Monuments Used to Define the Version 1.1 Transformation Surface**

| Mapsheet | Number of Monuments | Mean Difference (m) | 90-th Percentile (m) | Maximum Difference (m) | Standard Deviation (m) |
|---|---|---|---|---|---|
| 82/E | 314 | 0.43 | 0.93 | 3.01 | 0.15 |
| 82/F | 455 | 0.40 | 0.89 | 3.52 | 0.19 |
| 82/G,J | 229 | 0.71 | 1.50 | 4.79 | 0.49 |
| 82/K | 176 | 0.51 | 1.15 | 3.47 | 0.34 |
| 82/L | 178 | 0.25 | 0.53 | 5.84 | 0.23 |
| 82/M | 44 | 0.20 | 0.39 | 0.49 | 0.02 |
| 82/N,O | 30 | 0.99 | 2.95 | 3.75 | 1.46 |
| 83 | 116 | 0.65 | 1.75 | 3.36 | 0.48 |
| 92/B | 207 | 0.63 | 1.23 | 3.35 | 0.31 |
| 92/C | 87 | 0.30 | 0.62 | 1.15 | 0.07 |
| 92/E | 128 | 0.61 | 1.58 | 4.33 | 0.58 |
| 92/F | 523 | 0.53 | 1.05 | 9.54 | 0.38 |
| 92/G | 256 | 0.39 | 1.01 | 2.49 | 0.18 |
| 92/H | 283 | 0.21 | 0.40 | 1.14 | 0.04 |
| 92/I | 145 | 0.23 | 0.47 | 1.10 | 0.04 |
| 92/J | 104 | 0.33 | 0.61 | 1.50 | 0.07 |
| 92/K | 102 | 0.39 | 0.80 | 2.01 | 0.15 |
| 92/L | 130 | 0.45 | 0.82 | 3.15 | 0.28 |
| 92/M | 8 | 0.14 | N/A | 0.31 | 0.02 |
| 92/N | 164 | 0.30 | 0.60 | 2.35 | 0.09 |
| 92/O | 254 | 0.48 | 0.92 | 4.05 | 0.22 |
| 92/P | 65 | 0.41 | 0.80 | 1.95 | 0.12 |
| 93/A | 314 | 0.63 | 1.31 | 2.38 | 0.21 |
| 93/B | 72 | 0.40 | 0.87 | 1.22 | 0.10 |
| 93/C | 28 | 0.94 | 1.94 | 3.14 | 0.56 |
| 93/D | 28 | 0.99 | 1.95 | 2.52 | 0.57 |
| 93/E | 86 | 0.24 | 0.50 | 3.31 | 0.19 |
| 93/F | 42 | 0.27 | 0.68 | 1.48 | 0.10 |
| 93/G | 153 | 0.29 | 0.74 | 2.12 | 0.12 |
| 93/H | 28 | 0.38 | 0.92 | 1.09 | 0.10 |
| 93/I | 147 | 0.44 | 1.04 | 3.10 | 0.25 |
| 93/J | 80 | 0.23 | 0.46 | 1.52 | 0.08 |
| 93/K | 214 | 0.33 | 0.87 | 2.18 | 0.12 |
| 93/L | 269 | 0.40 | 0.98 | 3.64 | 0.19 |
| 93/M | 153 | 0.61 | 1.67 | 3.96 | 0.52 |
| 93/N | 63 | 0.71 | 1.83 | 2.86 | 0.49 |
| 93/O | 65 | 0.32 | 0.85 | 3.14 | 0.24 |
| 93/P | 92 | 0.24 | 0.44 | 2.27 | 0.12 |

Continued

| Mapsheet | Number of Monuments | Mean Difference | 90-th Percentile | Maximum Difference | Standard Deviation |
|---|---|---|---|---|---|
| 94/A | 53 | 0.47 | 1.13 | 2.08 | 0.18 |
| 94/B | 83 | 0.59 | 1.18 | 2.41 | 0.20 |
| 94/C | 64 | 0.72 | 1.46 | 2.50 | 0.29 |
| 94/D | 69 | 0.58 | 1.24 | 3.34 | 0.31 |
| 94/E | 80 | 0.48 | 0.81 | 1.82 | 0.11 |
| 94/F | 72 | 0.66 | 1.44 | 2.43 | 0.28 |
| 94/G | 68 | 0.58 | 1.28 | 2.72 | 0.27 |
| 94/H | 47 | 1.53 | 1.66 | 16.45 | 3.61 |
| 94/I | 181 | 0.69 | 1.43 | 10.30 | 0.89 |
| 94/J | 145 | 0.67 | 1.40 | 4.32 | 0.41 |
| 94/K | 55 | 0.24 | 0.60 | 0.91 | 0.05 |
| 94/L | 139 | 0.46 | 0.92 | 3.17 | 0.19 |
| 94/M | 86 | 0.57 | 1.21 | 3.12 | 0.34 |
| 94/N | 51 | 1.37 | 3.37 | 8.45 | 2.80 |
| 94/O | 177 | 1.37 | 2.75 | 8.52 | 1.99 |
| 94/P | 226 | 0.84 | 2.13 | 7.90 | 1.43 |
| 103/A,H, & E½ G | 34 | 0.40 | 0.78 | 2.82 | 0.25 |
| 103/B,C | 32 | 0.38 | 0.86 | 1.55 | 0.13 |
| 103/F,K, & W½ G | 124 | 0.61 | 1.37 | 3.03 | 0.35 |
| 103/I,J | 221 | 0.52 | 1.14 | 4.34 | 0.29 |
| 103/P | 68 | 0.67 | 1.43 | 3.74 | 0.44 |
| 104/A | 171 | 0.50 | 0.95 | 5.93 | 0.37 |
| 104/B,C | 80 | 0.52 | 0.68 | 20.92 | 5.41 |
| 104/F,G | 234 | 0.43 | 0.91 | 1.95 | 0.13 |
| 104/H | 96 | 0.84 | 1.93 | 3.75 | 0.63 |
| 104/I | 70 | 0.51 | 1.31 | 3.72 | 0.45 |
| 104/J | 136 | 0.80 | 1.39 | 2.64 | 0.26 |
| 104/K | 62 | 0.50 | 0.85 | 4.00 | 0.38 |
| 104/M | 87 | 0.22 | 0.55 | 1.55 | 0.08 |
| 104/N | 179 | 0.50 | 1.11 | 2.68 | 0.19 |
| 104/O | 98 | 0.32 | 0.62 | 2.39 | 0.12 |
| 104/P | 143 | 0.38 | 0.90 | 1.74 | 0.13 |

"INTTAB" routine available in the National Transformation Software package should be used to get a hard copy listing of the grid shifts in the region of interest. Ifthe grid shifts changes are reasonably smooth, the internal accuracy of the network will not be drastically affected. Large changes in the magnitude of the grid shifts will affect the internal accuracy of the network to be transformed.

## Independent evaluation of transformation

To independently verify the precision and reliability of the National Transformation and to ensure its efficiency and effectiveness in a production environment, the following tests were performed.

### 5.1 4th order control monuments test

Table 2 lists the statistics derived from the Fourth Order Monuments for which coordinates are available in both systems. These monuments were excluded from the definition of the Version 1.1 Transformation due to the questionable accuracy implied by fourth order. As well, they were set aside as reserve points for checking the National Transformation. The data set includes all Boundary Monuments, MCE and NAT.DEF. and DTS monuments that were available. These 4 th order monu-

**Table 2: Coordinate difference statistics for 4th order monuments, listed by quadsheet**

| Quadsheet | Number of Monuments | Mean Difference (m) | 90-th Percentile (m) | Maximum Difference (m) | Standard Deviation (m) |
|---|---|---|---|---|---|
| 82 | 274 | 3.09 | 10.07 | 46.51 | 5.42 |
| 83 | 38 | 12.44 | 25.49 | 77.29 | 14.88 |
| 92 | 548 | 2.02 | 4.32 | 68.93 | 4.54 |
| 93 | 375 | 2.08 | 4.37 | 35.26 | 3.91 |
| 94 | 40 | 5.23 | 13.37 | 22.55 | 5.67 |
| 103 | 52 | 2.70 | 7.27 | 9.56 | 2.73 |
| 104 | 112 | 3.15 | 8.65 | 22.60 | 3.65 |

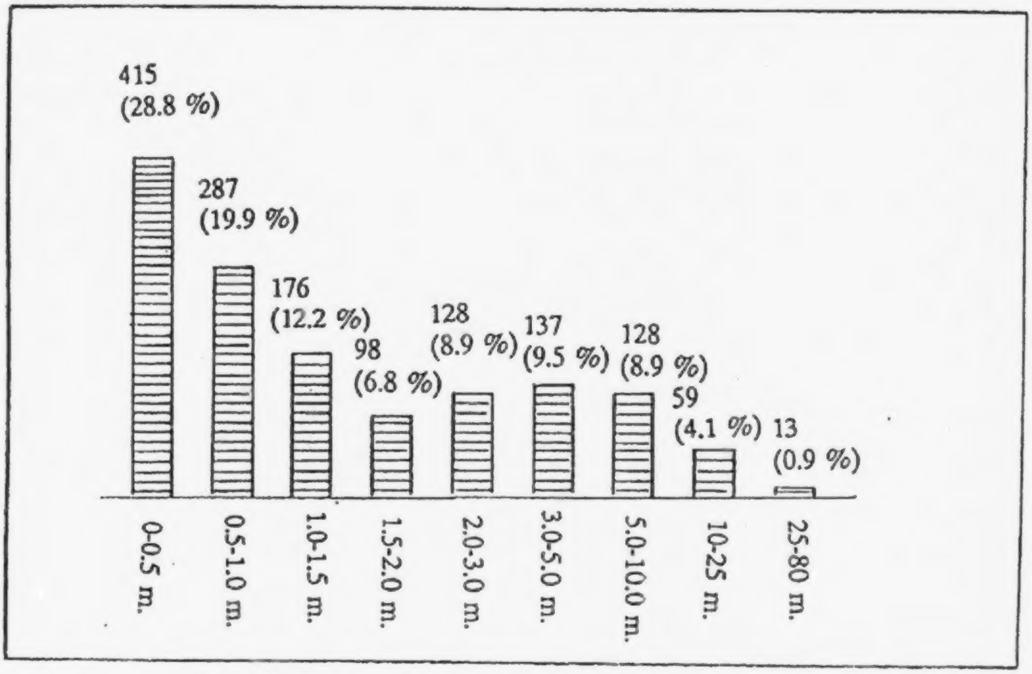


**Figure 11:** *Bar chart of coordinate difference statistics for 4th order monuments*

ments extend across the province. Figure 11 is a provincial summary of Table 2 in bar chart form.

The statistics are higher for this data set partly due to the lower accuracy of the NAD27 coordinates. The mean and 90-th percentile for the small corner of quadsheet 83 that falls within the province of British Columbia are high, thus special care should be exercised in this area. Nevertheless, on a provincial scale the results are very good. *95 % of the differences between transformed and rigorously adjusted coordinates are within 10 metres, while 68% of these differences are within 2 metres* (Figure 11).

### 5.2 Comparison with the NADCON transformation from USNGS

The software put out by the United States National Geodetic Survey (USNGS) for transforming coordinates from NAD27 to NAD83 (NADCON) within the continental United States has a limited region of overlap with that of the National Transformation. This section describes a comparison of the accuracy of the two transformation packages.

A test set consisting of 2558 monuments within the region of B.C. below 50° Latitude and east of 126° longitude was used. Monuments excluded from the definition of the transformation surface for the National Transformation were not rejected from this data set. Two monuments were excluded as obvious outliers with coordinate differences greater than 50 metres. Statistics of the coordinate differences between the two predicted sets and the NAD83 coordinates from the July 1990 readjustment are shown in Table 3, below.

There appears to be a systematic difference between the two solutions of 1 metre in latitude and 1.5 metres in longitude. Overall, the National transformation is providing transformations with better accuracy in the region of interest. Users working in this area with coordinates obtained from both transformation packages should be aware of this incompatibility.

### 5.3 Cadastral map data conversion pilot

When using the National Transformation to convert digital map data, usually an interface specific to the map system environment will also have to be built. Thus, an Intergraph based system will require reformatting of the Intergraph Design (IGDS) files so as to accept the National Transformation software. As well, the final file converted onto NAD83 will again have to be reformatted into the proper IGDS file format.

Currently, a pilot project within the Ministry of Lands and Parks, Surveyor General Branch, is examining the issues associated with converting the 1:20 000 scale cadastral map data using the National Transformation. Interface software was first written to accommodate the Intergraph IGDS file structure. Preliminary data within the four selected sites in the province (Saanich peninsula, Kootenays, Skeena and Peace River) show excellent results. The areas selected attempted to reflect a cross-section of data densities, age of data compilation and area of province. The quality control (map comparison) was provided by the NAD83 TRIM mapsheets, agreement with which so far is extremely good. The information converted was land tenure, planimetry and limited forestry information. *The largest file conversion in the Saanich peninsula (including downtown Victoria) entailed 138 071 coordinate conversions and took fifteen minutes in total elapsed time on a DIGITAL VAX 3500 computer.*

Final editing - entailing removal of old neat lines and replacing with new NAD83 surround;merging of surrounding affected files; continuity check at former map neat lines; and substitution of old planimetry with

**Table 3: Comparison between the National Transformation and NADCON**

| | National Transformation | | NADCON | |
|---|---|---|---|---|
| | $\Delta \varphi$ | $\Delta \lambda$ | $\Delta \varphi$ | $\Delta \lambda$ |
| Max (m) | 7.564 | 7.609 | 15.206 | 11.843 |
| Min (m) | -12.673 | -11.612 | -15.205 | -10.262 |
| Mean (m) | -0.012 | -0.073 | 1.067 | 1.509 |
| $\sigma$ (m) | 0.716 | 0.671 | 2.652 | 1.200 |

newer TRIM planimetry - amounted to the most time consuming part of the conversion exercise. For the mapsheets in this project, *the final editing averaged 3 hours per mapsheet.*

### 5.4 Thin plate spline function alternative

Quite apart from the tests reported here, an investigation by Barrodale Computing into thin plate spline functions for defining the transformation surface has also proved valuable in verifying the integrity of the National Transformation. The thin plate spline function is different from a polynomial fitting process, such as that used in the National Transformation, in that it fits the transformation surface exactly at the defining control monuments. While still under investigation by Dr. Ian Barrodale, the methodology may be well suited for the urban transformation surface definition later on since these(urban ) areas are thought to have fairly homogeneous and consistent control networks. That is, the least squares best fitting process of the polynomial at the control points, needed to reject non-conforming control networks/monuments, may be less accurate than an exact fitting thin plate spline in those situations.

Currently, the Ministry of Forests is evaluating the feasibility of the thin plate spline approach for the conversion of 1:20 000 forest cover maps. Since these maps represent a large investment it is desirable to convert them to NAD83 and overlay them against the newer TRIM (non-forest cover) base maps. However, the forestry maps were compiled using approximate methods for georeferencing so that their relationship with the geodetic reference system is considered tenuous. Hence the consideration of the thin plate spline function to deal with ***heterogeneous*** data sets.

## Transformations in urban areas

Further to section 3.8, *the current plans are to provide a new version of the National Transformation on April 1, 1993, to accommodate the transformation needs within ISAs.* The high density control networks, with their monuments spacings of 300 - 800 m., should permit a transformation surface definition that will yield 5 - 20 cm. accuracy, depending on density and networks consistency and accuracy. *This should be generally sufficient to transform map bases of 1:1 000 and smaller.*

Users requiring NAD83 ISA coordinates immediately may approach Surveys and Resource Mapping Branch, Geodetic Control Unit, to arrange for preliminary information. *Preliminary coordinates will not change by more than 0.5 m. when the final values are published in 1993.* As well, the relative accuracies between control monuments will not change significantly (say, within 3rd Order accuracy). Therefore, municipalities wishing to proceed on NAD83 now may be able to easily perform up to 0.5 m. shifts of their NAD83 data holdings in 1993, without significant loss of accuracy. Of course *each organization should study its requirements carefully* in order to decide whether these small errors (few cm.) after the shift affect their application(s).

## National transformation software and users guide

The actual shifts from NAD27 to NAD83 for any given point(s) in B.C. is computed using the National Transformation software. The Users Guide for the National Transformation software is published within the CISM (1990) publication 'Proceedings of the CISM seminar; Moving to NAD83, the new address for georeferenced data in Canada'. It is also available with the *National Transformation software as distributed by MAPS-BC at Surveys and Resource Mapping Branch.*

The diskettes supplied from MAPS-BC contain the source and executable codes for programs GRIDPT, DIRFIL, INTGRID, INTTAB, READDA and GSRUG. The source code is compatible with DIGITAL VAX, Macintosh,and IBM/PC computer platforms. There are two versions of the executables supplied. The ones with a 7 at the end of the name require a math co-processor, whereas the other ones do not require a math co-processor. As well, the diskettes contain the actual grid data required by the software execution. The data only covers the British Columbia land mass and an arbitrary buffer around it.

For assistance with the National Transformation in B.C., contact the Geodetic Control Unit at Surveys and Resource Mapping Branch, Ministry of Lands and Parks, in Victoria.

## Summary and conclusions

The analysis of the 9272 First to Third Order monuments used to define the transformation surface indicates a mean difference between the transformed and adjusted NAD83 coordinates of half a metre. Ninety percent of the monuments exhibited coordinate differences of less than 1.14 metres. Actual values can vary significantly throughout the province. The transformation should provide coordinate conversions with an accuracy acceptable to many applications such as GIS, cartography and thematic mapping. Users with much higher accuracy requirements may need to look at other alternatives.

The statistics presented in this paper are meant as a guideline only. They give a general idea of the accuracy of the transformation from NAD27 to NAD83 for the monuments utilized for this analysis. Positional data maintained by other agencies may be of far differing accuracy and internal reliability and will thus produce widely varying results. In addition, the statistics presented here cover broad regions. Smaller, more localized regions can have significant differences from the mean values.

To offer practical guidance, conversion to NAD83 of existing NAD27 1:10 000 scale mapping (and smaller scale) will be supported by this National Transformation version 1.1. Nevertheless, it is recommended that any agency contemplating transforming their data holdings complete further studies, if possible, to identify areas of concern — especially when transforming larger scale mapping. Of particular interest will be positional data sets tied to monuments excluded from the National Transformation surface definition. Entire networks which were found to be inconsistent with the surrounding control were excluded from the definition of the transformation surface. Many of these networks can be viewed as localized datums. A transformation specific to these monuments can be derived that will provide far superior accuracy in the transformed coordinates but the compatibility with the national standard will be lost. A good source of quality control in checking a conversion is the NAD83 based TRIM 1:20 000 mapping. However, this mapping is incomplete at present, covering 25 % of the province (100 % by 1997).

Finally, users should exercise extreme caution in using version 1.1 in developed (urban) areas of the province since the rigorous readjustment of the geodetic control networks onto NAD83 is as yet incomplete. A new version of the transformation based on the new provincial readjustment and incorporating the Integrated Survey Areas (i.e. many of the urban municipalities in the province) will be made available April 1, 1993. It is expected that this new version will support 1:1 000 scale (and smaller) mapping conversions in these ISAs, while continuing to support 1:10 000 scale (and smaller) mapping conversions in other areas of the province. Note that the new version will not change significantly outside of the ISAs, i.e. the changes will be well within the tolerable limits of the largest map scales (1:10 000).

The cadastral data conversion pilot conducted by the Surveyor General has shown the viability of the National Transformation in a production setting, in this case conversion of 1:20 000 cadastral map data residing in Intergraph's IGDS file format. Thus, entire map files are converted in two to fifteen minutes, although final editing requires about 3 hours per map.

The NAD83 technical challenge now lies largely with the mapping and LIS / GIS communities (as opposed to the geodesists). While the conversion of point coordinates from NAD27 to NAD83 is easily achieved with the National Transformation, the treatment of map topology and integrity will vary from one application to another, due to both the hardware / software system environment and the information content characteristics. The greater challenge of course is to have sharable and seamless data structures built, while accommodating NAD27 to NAD83 transformed data.

## Acknowledgments

My thanks go to the CCSC, the Canadian Council on Geomatics (CCOG) and especially EMR Canada's Canada Centre for Surveying (CCS) for the national direction and coordination of the NAD83 project, and for providing excellent support in developing a workable version 1.1 of the National Transformation in British Columbia. Particular thanks go to Mike Pinch, Paul Henderson, Don Junkins, Steve Farley, Dan O'Kane and Gord Garrard of CCS in this regard. Don Junkins was instrumental in formulating the National Transformation methodology.

I would like to take this opportunity to acknowledge the efforts of staff within my Unit for the dedicated and thoughtful work that continues to be applied within the NAD83 project. Ted Rowlandson and his group are responsible for the data preparation and the actual readjustment exercise, whereas Kent Pointon (no longer with the Branch ) provided significant contribution to the National Transformation analysis, the results of which are presented in this paper.

* *Kent Pointon is now President of Manthanein Geomatics Ltd. in Vivtoria, BC*

## References

Balser, R.W. (1990) *The Implementation of NAD83 in British Columbia - Impact on Mapping Programs and Implications for Data Management,* Proceedings of the CISM Seminar; Moving to NAD'83, the New Address for Georeferenced Data in Canada, Ottawa

Boal, D. and Henderson, P. (1988) *The NAD83 Project - Status and Background,* Papers for the CISM Seminars on the NAD83 redefinition in Canada and the impact on Users; CISM, Ottawa.

Dewhurst, W.A. (1990) *NADCON The Application of Minimum Curvature-Derived Surfaces in the Transformation of Positional Data From the North American Datum of 1927 to the North American Datum of 1983,* NOAA Technical Memorandum NOS NGS-50, U.S. Dept. Comm.

Duffy, D. A. (1990) *Land Information Management The Provincial Strategy as it relates to Local Government,* proceedings of the GIS'90 Symposium, March, 1990, Vancouver

Farley, S.A. and D.J. Junkins (1990) *User Guide for the National Transformation Interpolation Software and NAD27-NAD83 Data Set,* Proceedings of the CISM Seminar; Moving to NAD'83, the New Address for Georeferenced Data in Canada, Ottawa

Henderson, P. et al (1989) *The NAD83 Project - Background and Status*, Canada Centre for Surveying, Ottawa

Junkins, D.J. (1989) *Notes on Preliminary Definition of a National Transformation*, Minutes of the February, 1989 Meeting of the Canadian Control Survey Committee.

Junkins, D.J. (1990) *The National Transformation for Converting Between NAD27 and NAD83 in Canada*, Proceedings of the CISM Seminar; Moving to NAD'83, the New Address for Georeferenced Data in Canada, Ottawa.

Junkins, D.J. (1988) *Transforming to NAD83*, Papers for the CISM Seminars on the NAD'83 Redefinition In Canada and the Impact On Users, CISM, 94-105.

Kassam, A. and Balser, R.W. (1990) *Implementation of NAD83 in British Columbia*, Proceedings of the CISM Seminar; Moving to NAD'83, the New Address for Georeferenced Data in Canada, Ottawa

Kassam, A. (1989) *New Alberta Survey Control System, MASCOT*, Proceedings, Challenge for the 1990's GIS, February 27 - March 3, 1989; Energy Mines and Resources Canada and The Canadian Institute of Surveying and Mapping, Ottawa.

Kenk, E. and Hofmeyr, H. A (1990) *Land Information Infrastructure: Government of British Columbia Perspective*, paper for Urban and Regional Information Systems Association (URISA) 1990 conference in Edmonton, August 12-16, 1990.

MAPS-BC product (1991) *NAD27 to NAD83 conversion in British Columbia: National Transformation Version 1.1*, Surveys and Resource Mapping Branch, Victoria

O'Brien, L.J. (1985)*Changing the Shape of Canada*, GEOS, 1985-4; Energy,Mines and Resources Canada, Ottawa.

Pinch, M.C. (1990) *Differences Between NAD27 and NAD83*, Proceedings of the CISM Seminar; Moving to NAD'83, the New Address for Georeferenced Data in Canada.

University of Maine at Orono (1984) *The use and value of a Geodetic Reference System*, Research study by The University of Maine at Orono (1984), Orono, Maine 04469

Wilkinson, G.A. (1988) *NAD83 - Status and Plans in British Columbia*, Papers for the CISM Seminars on the NAD'83 Redefinition In Canada and the Impact On Users, CISM, 134-139.

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Putting the records in order: GIS database construction

**C. Warren Ferguson**

Cantex
1425 Pearson Place
Kamloops, BC
Canada, V1S 1J9

## Abstract

*The construction of an accurate, intelligent, and above all, usable database is of utmost importance in the development and implementation of a Geographic Information System. This paper will begin by looking at the essential elements of the GIS database construction process. Next, it will explore the requirements for an accurate database construction process, comment about new technology and upcoming improvements and, lastly, share a few ways to achieve success in geographic database construction. Some specific project experiences will be shared.*

## Introduction

After a brief introductory history of GIS conversion and why it is implemented, I'd like to touch on the essential elements of the GIS database construction process. Next, we'll look at the requirements for an accurate database construction process, add a few comments about new technology and upcoming improvements, and lastly, share what I think are some ways to achieve success in geographic database construction. As we proceed, I will share some specific project experiences.

The conversion process as we described it in 1980 had seven steps to it. Let me review those. First, landbases had to be created. Second, the data we wanted to add had to be corrected because it was rarely accurate in the form that we found it. One of the things that we did then and still have to do today is standardize the records in preparation for entering them into the computer database. We had to establish data relationships because in many cases they were not evident from the manual records and the maps we had. Then the facilities had to be digitized.

After all the information was in the computer, there was considerable time spent doing edits to insure that we had the accuracy and the correctness we wanted. Most often we found that there were several correction cycles required in order to get the data in the format that we finally wanted. Even back in the early '80s it became obvious to me that the trademarks of excellence in conversion were accuracy, low cost and timeliness. These are items that were easy to identify with success and they still are the result of excellence in conversion.

In the early '80s conversion was a mystery and many people were looking for that magic box where they could stuff their original in on one side and get a digital tape out the other side. Even then, there were those who wanted it yesterday. There was no way to satisfy all of their many demands.

Conversion, or as you will hear me suggest, the activity better described as geographic database construction, is a very unique process and one that John Antenucci has recently called "the worst of both worlds because it is a process that is both capital intensive and labor intensive". I don't particularly agree with John's description that it is the worst of both worlds. But I will tell you that it is capital intensive and it is my conviction that people are what make success in conversion.

Let's back up a minute and look at why you have chosen to implement a GIS system. Most often it's a combination of several factors. There probably are problems with your existing manual system. Perhaps more efficient management of your information is the need in your organization. Surely all of us can be more cost effective. In today's times and with the need for infrastructure renewal, we find that we must have an accurate inventory of our facilities.

When we look at putting the records in order, one of the things we must remember is that the problems with a manual system is one of the most compelling reasons to develop a GIS project. Very often there are several things that are found lacking in existing manual systems. First is the physical condition of the documents. They're simply old, they're tattered, they're torn, they have coffee stains on them, they may be very difficult to read, and in some cases they may have even been lost. A proper examination often reveals that there is a very high cost to maintain these paper documents. Sometimes there is confusion over which set of records are the "official" records. Often there is duplication of records, and obviously duplication is a needless waste of time, energy, and money. Very often, the important records are in the heads of those old-time employees who know exactly where everything is and what keeps the system working.

These dilemmas usually don't show on any of the records. Added to this is the difficulty of keeping records current. To some extent this is the description of the records that we have to work with when we try to put the records in order and do data conversion. But the manual record is not the only thing we work with these days when we begin conversion. In the mid-to-late '80s we recognized that conversion was the process for building a digital facilities model on a geographic base from mechanized records as well as existing manual records.

## The database construction process

We talk about database construction as a process with four components. Four essential components. The first, information sources, prompts us to ask the question, what are the inputs to the process? Second, the technology causes us to define what tools are going to be used to perform the process. Third, methodology makes us ask how the technology is applied to perform the process. The fourth essential component is the output or end product. How sophisticated must it be? How much is expected of the outputs from the process?

Again, conversion is a process, it is not an event. It will not take place suddenly, quickly, or all at one time. It's a lot more than most first imagine.

The conversion process involves much more than just converting the existing data to digital format. It involves an accurate landbase created to meet present needs as well as needs anticipated for the future. It represents a standardizing of symbology throughout our record set or database. The data sources for input have to be compiled and methods found to combine the variety of sources that exist within most organizations. Some data may already be in a form of mechanized records. Finally, it involves converting, changing, modifying, adding and verifying all existing and acquired data to build the geographic database.

## Requirements for an accurate process

The requirements for quality information in a geographic database are paramount. The data converted must have the following characteristics: It must be accurate, and that means free from mistakes and errors; whatever you see is correct. It must be complete, and that means that all necessary elements are displayed appropriately. It must be timely, current and up to date. It must be reliable if users are to depend on the information. It certainly must be credible, which means that different output products from the common database must agree when viewed or used by a variety of users. In other words, different sources must agree. The information must pass validity checks because these are necessary if the impact of changes is to be predictable (and it must be). Finally, the information must be convenient and readable in that it is clean, legible, easy to use, and easy to read.

At this point let me say that to achieve the excellence in conversion that I believe is appropriate, each conversion process must be tailored to fit the user organization perfectly. That's why each conversion process is different unto itself, driven by the needs of the specific organization, the users themselves and the existing or available data that can be included in the geographic information system database.

As we move forward we see that data conversion involves a variety of data sources, both paper records and electronic records, as well as data captured in the field. Certainly the quality of records has a big impact on our ability to create an appropriate database. Many people have to make a strategic decision with respect to the alternatives of capturing historical data or collecting data from today forward. As geographic information systems become more pervasive throughout our organizations there is an increasing expectation for accuracy and demand for quality. These can only be achieved through appropriate quality control measures in the conversion process and a quality assurance procedure performed upon completion of the database.

In recent months, I have come to feel that geographic database construction is a more appropriate description of the process than is conversion. Clearly, to build a geographic database requires a lot more than simply conversion as data entry. It involves the design of real world data models so that the applications we have today, as well as those that will step forward in the future, can be enabled by our database. Very often we have to inventory or acquire data. Augmenting the existing data is nearly always a requirement of the geographic database construction process. Digitizing may be either conventional in nature at a digitizing table or scan digitizing and vectorizing those records into the digital database. A key activity in nearly all of the work done today is to translate and merge data from other digital files already existent within the user organization. Often there is a requirement to blend multiple databases to achieve the best of the variety of source data that is available.

Finally there is an absolute requirement to create connectivity throughout the geographic system database. This connectivity is a mandatory requirement of both linear and polygonal networks that form the topological basis of our geographic information systems. As you know, topology is generated by line/node structures. Nodes can be formed at the intersection between lines. They define the connectivity between adjacent line segments. Nodes may have associated user defined attributes as well. In linear networks lines are used to represent features such as roads, pipes or cables. Nodes represent the point at which these features connect to each other, for example a valve or street intersection.

In polygonal networks lines define the boundaries of areas and nodes represent the points where shared boundary lines meet. Usually an identifier for the enclosed areas is included to define topological relationships more thoroughly. An example of a polygonal network might be the service area boundary of a particular district office.

The skill sets required for GIS database construction are certainly far more than just data entry. Most projects today involve perhaps geodesy and/or photogrammetry. For sure, however, they require computer science skills, digital cartography, and a variety of engineering skills. In addition, database design and network understanding are required skills for the construction of a good, useful GIS database. The age old skills of project implementation and quality assurance are certainly required to build the complex GIS databases that we use today.

If we look at it in light of today's technologies, today's GIS systems and the GIS database construction requirements, we see that what we're talking about is making a "real world database of geographic information". The information that we put in this database tends to be landbase information which is geographic and is related through specific land data elements as well as the coordinate system. We include in our database a great deal of boundary information which is polygonal in format. And finally we put in asset or facility information, generally speaking the location, the relational features of that asset, as well as descriptive attribute information.

Why all the talk about conversion? What's the big deal? Money! If you recall the typical system development cycle for a data processing system you remember how little conversion costs — maybe 15%. But in the GIS system development cycle conversion can cost 50% to 60% of the total. And some people say it can cost as much as 75%, but that's only if you already own the hardware and system software.

When you compare GIS system development to a conventional data processing system, you see how much different the conversion cost is (as a percentage of the total system development). However, you should remember that while hardware has a 5-8 year life cycle and software a 3-5 year cycle, the data has a 25-40 year life cycle.

What makes up the cost of GIS database construction? To many it's a puzzle, but to those of us that do it for a living there are a wide variety of factors that influence each specific project. The base maps that are to be used for the landbase, the geographic accuracy, the facility density, the geographic size of the organization, as well as the scope and complexity of the project are those that most often impact cost on a large scale.

There are several development and conversion alternatives. Do it yourself, hire a consultant, purchase software, hire a conversion company or combinations of all these. Most organizations use some combination, or series of combinations, to gain the greatest benefit.

Let's be sure we know what are the functional requirements. Conversion costs a lot and we must do it right the first time. Functional requirements are answers to the what, when, and why questions (not the how question - that comes later). We must determine the project scope, because scope affects the requirements. Is this a corporate project or only a district project — do we want full functionality from the start or must we more quickly solve some onerous problem?

We must also look at our present system and ask what should be Changed? Deleted? Streamlined? Added? "What if" and "if then" analysis must be evaluated vs benefits. And the result stated as a functional requirement. When we look at all the functional requirements we see the magnitude of the risk and the challenges. GIS cuts across departmental lines. We must solicit requirements from the right levels within the organization. Don't assume the current procedure/organization is correct. Try to identify new opportunities.

Nearly all users decide to use a conversion service company because the cost is reduced, the accuracy is improved by the conversion specialist's verification or QA software, and the data is converted much more quickly.

A clear scope of work is simply a must. It must include:

- Geographic Coverage
- Feature types and quantities
- Number/type of customers
- Existing records - types, quantities
- Project subdivisions
- New records - types, quantities
- Pilot project
- Backlog handling

The schedule must be defined and resources allocated.

Existing data specifications such as types of data, drawings, file records, computerized data, graphic specifications, methods and procedures documentation, computer record formats, and field inventory notes are required.

Deliverables must be clearly defined. Physical format — computer media or hard copy — plots, aperture cards, database reports, edit reports, or all of these must be specified.

The host-dependent GIS data structure — logical/physical file structures, working units, naming conventions, symbology, feature definitions, mapping standards, positional accuracy, and drafting standards must be clearly defined.

Database specifications — data structure (host/dependent), entity/attribute definition, data relationship and data sources, plus priorities of multiple sources, as well as valid attribute values/default values and critical/non-critical attributes must be defined.

With this we get to the point of decision making — are we going to be in the action, or are we just going to tell stories about the action? It takes a knowledgeable and a committed executive to decide to proceed.

## Technology innovations

Innovations in conversion technology can improve not only the cost of converting to a GIS system, but the accuracy and timeliness of the conversion effort as well. Innovations come about in two ways. There are some that have great benefit within an existing functional area of the conversion process, and others that use new technology to redefine the conversion process itself.

An example of the first type of innovation (one that benefits an existing functional area) might be the use of artificial intelligence (AI) in automated editors to perform data purification and edits. An example of a new technology that essentially redefines the conversion process might be the use of scanning with pattern recognition and intelligent object association combined with record class analysis and clean-up. This could create an automated conversion process.

**Table 1:** Identifies various technologies that may either improve the typical conversion process, redefine the conversion process, or both.

**Table 1**

| **Technology** | **Improve typical Conversion process** | **Redefine the Conversion process** |
|---|---|---|
| Artificial intelligence | X | X |
| Scanning & vectorization | X | X |
| Voice data input | X | |
| Pattern recognition | | X |
| Neural networks | | X |
| Automated edgematching | X | X |
| Statistical character recognition | | X |
| Faster workstation | X | |
| New CPU's | | X |

Since conversion is usually a large part of an GIS project's cost, there have been, and are now, several innovations being touted as "major breakthroughs". It is important to recognize that, as a new technology proceeds from introduction through design, expectations rise greatly. Then, as applications are implemented, and as the new technology is integrated into the process and production use occurs, expectations are reduced to a practical value or proven reality (that may or may not be an improvement).

The conversion industry has seen this pattern occur frequently in the recent past. Innovations to date have been limited in their ability to reduce conversion cost. Most of the time, results have been lower than expectations. Nonetheless, the state-of-the-art in GIS data conversion has advanced dramatically in the past five years. Scanning and vectorizing is cost effective if the maps to be converted are up-to-date, complete, and of reasonable quality. There now exists a great deal of specialized conversion software that enables conversion operations to take place much faster than can be accomplished with system vendor software. Translation programs now exist that enable data to be converted on one system and translated to the user's system.

While these are all welcome improvements in conversion technology, they have reduced the cost of conversion only enough to maintain its level at about 50% of the total cost of GIS implementation. This is because CPU, workstation, and GIS software costs have also decreased substantially in the past five years.

Conversion technology is clearly improvable. It is taken for granted that tomorrow's technology will be far more powerful than today's. But people must grow and improve as well, or the innovative technology will not mean much. Conventional wisdom tends to blame "the computer" for the foul up, but results rarely have much to do with the mechanism that executes the underlying math or logic. Instead, the credit or blame should go to the people who understand, or fail to understand, the problem they assign to a particular information processing system.

Let's return to a perspective of where we are today in the world of computing. In the '50s computing was number crunching. Data, and data management, was the focus from the early '60s until the late '70s. The Information Age began in the late '70s. At precisely the same time GIS emerged as a unique information technology.

Looking forward we can see that very quickly our corporate information will become knowledge based. It will also be geography based. The application potentials are truly unlimited. The application of geographic knowledge based technology is an area we should all learn more about. I urge you to find out about it.

Let me take a moment to summarize.

## GIS database construction - a summary

- GIS Database Construction is a construction process, not a conversion event.
- The process must be tailored to fit the user perfectly.
- It is a large, complex, detail-oriented job.
- It is both capital and people intensive.
- Accuracy, low cost and timeliness are the trademarks of success.
- GIS Database Construction enables GIS benefits to begin.
- With the records in order you have a solution that turns your data into profits.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Technology trends for GIS applications

**Michael J. Fisher**

Data General Corporation International
GIS Marketing Consultant
3400 Computer Drive
Westboro, MA USA 01580
Tel: 508-366-8911

## Abstract

*Both hardware and software development is under way, focused on improving performance and security issues realized when running database-critical applications, such as Geographic Information Systems (GIS). The marriage of multi-threaded relational databases and multi-processor based systems appears to be a trend in meeting the needs of the large GIS environment.*

*As a GIS database grows more sophisticated and increases in size, access to information and performing complex analysis create a burden on both CPUs and databases. When combined with the importance of protecting the most valuable aspect of a geographic information system - the data - these represent the most common problems for advanced GIS users.*

*This presentation offers insights into these developments by computer manufacturers and relational database vendors while outlining associated benefits. Industry standard benchmarks and other related specifications will be presented in support of the paper's findings.*

# D6 Scanning and conversion

---

**Automated map data conversion**

Wayne Coleman, Laser-Scan, Inc., Sterling, VA, USA

**Impacts of scanning resolution on GIS data conversion**

Lisa Klapwyk Crouser et al., Infotec Development, Inc., Portland, OR, USA

**Scanning technology: considerations in the preparation of scanner-ready manuscripts**

Daniel Haskell et al., Infotec Development, Inc., Portland, OR, USA

**Data collection for GIS applications using scanning in conjunction with automated raster to vector conversion technology**

Jeff Marceil, Scan Conversion Services, Burnaby, BC, Canada

---

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Automated map data conversion

**Wayne C. Coleman**

President
Laser-Scan
45635 Willow Pond Plaza
Sterling, VA, USA 22170-4457

## Abstract

*GIS are dependent on, as well as hampered by, the availability of digital data. Historically, GIS applications development has been dependent on the time-consuming and expensive process of manual digitizing. However, today there are numerous companies providing a wide range of cost-effective scanning systems that convert hard copy maps to raster data. Additionally, there are automated systems that efficiently convert raster data to intelligent vector map data.*

*It is generally agreed within the GIS community, that the cost of converting hard-copy source maps into intelligent attribute-coded vector data constitutes 60 to 80% of the total GIS expenditure. The requirement of converting data for mapping and GIS can vary in complexity from simple "spaghetti" linework to multi-attribute object level topologically structured data.*

*The success of emerging automated data conversion technology is measured in the final quality assurance phase. If the converted data is not tested and proven to meet all specifications and requirements, the conversion is a failed exercise.*

*Incorrectly converted and misrepresented data is worthless. For this reason the evaluation and procurement of automated conversion systems should be as well planned and as thorough as the procurement of the GIS system itself.*

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Impacts of scanning resolution on GIS data conversion

**Lisa Klapwyk Crouser**
Infotec Development, Inc.

**Dan Haskell**
Infotec Development, Inc.

**John Dabritz**
U.S. Forest Service, Pacific Northwest Region
Infotec Development, Inc.
500 N.E. Multnomah, Ste. 329
Portland, Oregon 97232, USA
Tel 503-238-0034
Fax 503-238-1224

## Abstract

*Scanning technology can provide the opportunity for efficiencies in GIS data conversion. However, the extent of these efficiencies is directly effected by the resolution of the scanned image. Map scale, data accuracy, line work quality, and data intensity influence choosing a scanning resolution. Resolution, in turn, impacts machine sizing, processing, and storage requirements. These cause-and-effect relationships have been evaluated using Line Trace Plus (LTPlus) data capture software at Infotec Development's GIS/LIS facility in Portland, OR. LTPlus was developed by the U.S. Forest Service in cooperation with the Soil Conservation Service. It operates under UNIX on a PC 386+ workstation. The "average" RAM and hard disk size of the workstation is in excess of 8Mb and 300 Mb respectively.*

## Introduction

The cost of converting data for use in a GIS is a concern for most users at some point in time. Scanning data for use in a GIS has always been an alternative but has become more affordable due in part to the increased processing power available in 386/486 PCs. These systems can provide users with the capability to convert the large data maps that natural resource analysis requires. Still, converting scan data requires a good size machine, and a budget conscious GIS user looking at scanning as a solution for data conversion needs to know what variables will impact system size, speed, and storage.

Scanning resolution is one of these variables. It is anticipated that scanning resolution impacts all phases of the data conversion process and carries over into the GIS analytical process. Validating its impacts on conversion processing and the final GIS output is the focus of this paper.

Infotec Development Inc. is a software engineering and integration firm that has been providing GIS/LIS ser-vices since 1981. The GIS/LIS facility in Portland, Oregon specializes in implementing GIS for natural resource management and has been providing GIS scan conversion services using Unix-based 386 systems since October, 1990.

## Determining scanning resolution

The choice of a scanning resolution is influenced by data accuracy, line work quality, data intensity, and map scale. The first objective in selecting a scanning resolution is to choose a high enough resolution that it will not degrade the positional accuracy of the data. When a raster representation of a line is vectorized, the resulting vector can deviate slightly to either side of the original line. Therefore it is important that the total distance a line "drifts" as the result of vectorization is not more than the on-the-ground-accuracy. Since the deviation is usually one pixel in width there is a general starting rule for determining a scanning resolution. The pixel size should be, at most, half of the on-the-ground accuracy of the original data.

Resolution can be calculated as follows

1. Determine the on-the-ground accuracy of the data

   Example: on-the-ground accuracy is 50' (15.24m).
2. Divide the on-the-ground accuracy in half to determine the maximum pixel size.

   Example 50' / 2 = 25' maximum pixel size required (15.24m / 2 = 7.62m).
3. Determine what distance is represented by a map inch (mm).

   Example 124 000 scale is equivalent to 1" = 2000' (1mm = 24m).
4. Select a dots per inch (dpi) scanning resolution

   Example 100 dpi (3.94 dpmm).
5. Divide the map inch distance (#3) by the scanning resolution to determine the size of the pixel.

   Example 2000' / 100 dpi = 20' per pixel (24m / 3.94 dpmm = 6.08m).
6. Compare pixel size with the quotient in #2. Pixel size should be less than the quotient.

   Example 20' pixel < 25' quotient from step #2 (6.08m < 7.62m).

Once this baseline is determined, the following factors may require increasing the scanning resolution (decreasing pixel size)

Density of data on the manuscript
Quality of linework
Number and tightness of curves
Number of low angle intersections

## Setting up the analysis

In order to examine how scanning resolution impacts file sizes, editing times, and end product quality it was necessary to define the project parameters. Infotec's GIS/LIS facility expertise is in the conversion of natural resource data. This data rarely has on-the-ground accuracy that exceeds USGS National Map Accuracy Standards. The accuracy standard for a USGS 124 000, 7.5' quadrangle is 40'.

Condition Class data was selected for the project. Condition Class maps contain productivity and size/age class information for major timber species and is representative of thematic information required in natural resource management. The Condition Class map was scan-ready and contained the following characteristics

- On-the-ground accuracy is 60' (18.29m)
- Drafted on a stable base media (mylar)
- Drafted using opaque black ink.
- 124 000 Map scale, USGS 7.5' quadrangle
- Approximately 210 map objects (1144 segments)
- Attributes on a separate overlay
- Scan input format is a binary Run Length Compression file (.RLC)
- Scan is monochrome
- Output format is an ASCII file, USGS Digital Line Graph 3 (DLG-3)

The analysis was performed on a 386/20 LTPlus Workstation. The workstation components are listed in Appendix A.

First the Condition Class data was scanned at four resolutions, 75, 100, 150, and 200 dpi, which corresponds to 2.95, 3.94, 5.90, and 7.87 dots per mm. Given that the on-the-ground accuracy of the data was estimated at 60', the four scan resolutions easily fit our primary criteria for determining a resolution. All four provide a pixel size that is less than half the on-the-ground accuracy (30'). The relationship of resolution to pixel size and on-the-ground accuracy and scale is represented in Tables 1-1 and 1-2.

Once the map had been scanned, it was processed using LTPlus software. LTPlus was developed by the U.S. Forest Service in cooperation with the Soil Conservation Service. LTPlus was not designed to be a GIS, rather it was designed as a GIS "Front-end" capable of converting data for use in a number of GIS packages.

In simplest terms, LTPlus begins with raster input and ends with topologically structured vector output. This output can then be imported into a GIS for database linking and analytical processing. The basic LTPlus conversion process used for the project is outlined below.

1. Set up the LTPlus map file
2. Gather the input data. LTPlus recognizes data from three main places
   - Importing a scan file
   - Manually digitizing (onto raster)
   - Adding data from an external vector source (onto raster)
3. Edit the raster
   - Pass one - separate touches

**Table 1-1** Relationship of resolution to pixel size and On-The-Ground Accuracy and Scale - U.S. equivalent conversion

| U.S. Equivalent Conversion | | | |
|---|---|---|---|
| Scale<br>1:24,000 | Dots per Inch | Feet per Pixel | Maximum on-the-ground accuracy represented |
| 1"=2000' | 75 | 26.67 | 54' |
| 1"=2000' | 100 | 20.00 | 40' |
| 1"=2000' | 150 | 13.33 | 27' |
| 1"=2000' | 200 | 10.00 | 20' |

**Table 1-2** Relationship of resolution to pixel size and On-The-Ground Accuracy and Scale - Metric System Conversion

| Metric System Conversion | | | |
|---|---|---|---|
| Scale<br>1:24,000 | Dots per mm | Meters per Pixel | Maximum on-the-ground accuracy represented |
| 1mm=24m | 2.95 | 8.14 | 16.3 m |
| 1mm=24m | 3.94 | 6.08 | 12.2 m |
| 1mm=24m | 5.90 | 4.07 | 8.2 m |
| 1mm=24m | 7.87 | 3.05 | 6.1 m |

4. Assemble to create vectors (topology) and quality control
   - may return to Step 3
5. Edgematch and snap data to coincident lines in external data
   - may return to Step 3 or 4
6. Attribute, attribute a second time, then compare for accuracy
7. Output

Integration of vector data from external sources, edgematching and attributing were not evaluated as they are primarily vector processes and their relationship to scan resolution is minimal.

## The effect of resolution on file size

To begin this project, the map data was scanned at standard E size (24"x30") even though the actual data only covered approximately 18"x24". This was done to establish the same scanning area at all four resolutions. The data was scanned into Run Length Compression (.RLC) format which compresses the data and reduces file size. Still, the high resolution 200 dpi scan file is over 2.5 times larger than the 75 dpi scan file. When the scan files were imported into LTPlus, and the data is decompressed for use, the 200 dpi LTPlus raster file is over 7 times larger than the 75 dpi LTPlus raster file. Table 2 shows the original scan and LTPlus raster sizes and LTPlus import times.

**Table 2.** Impacts of resolution on .RLC Scan and LTPlus raster Files

| Scanning Resolution | 75 dpi (2.95 dpmm) | 100 dpi (3.94 dpmm) | 150 dpi (5.90 dpmm) | 200 dpi (7.87 dpmm) |
|---|---|---|---|---|
| Resulting Pixel Size | 26.67' | 20.0' | 13.33' | 10.00' |
| .RLC Scan File size of an E size drawing (24"x30") | 0.165 mb | 0.219 mb | 0.329 mb | 0.440 mb |
| Comparison of .RLC Scan File Size (Ratio to 75 dpi) | 1 : 1 | 1.33 : 1 | 1.99 : 1 | 2.66 : 1 |
| LTPlus Import Time (Ratio to 75 dpi) | 1:1 | 1:2 | 1:4 | 1:6 |
| LTPlus Raster Size | 0.504 mb | 0.900 mb | 2.034 mb | 3.600 mb |
| Comparison of LTPlus Raster File Size (Ratio to 75 dpi) | 1 : 1 | 1.78 : 1 | 4.04 : 1 | 7.14 : 1 |

It is possible to reduce file size without altering the resolution of the scanned data. Some scanners allow the user to define or crop the map area to be scanned. Cropping a scan minimizes "null space" outside the data area and the result is a smaller scan file.

GIS scan conversion software, such as LTPlus, can further reduce raster file size. Once the data is in LTPlus, the user registers the raster using the grid tics of the scanned manuscript. After the map geodetics have been entered and the registration verified, the user can then "margin from the registered raster". Margining will remove any or all data outside of the registration. For example margin 20 erases all data that lies greater than 20 pixels outside the registration (Figure 1).

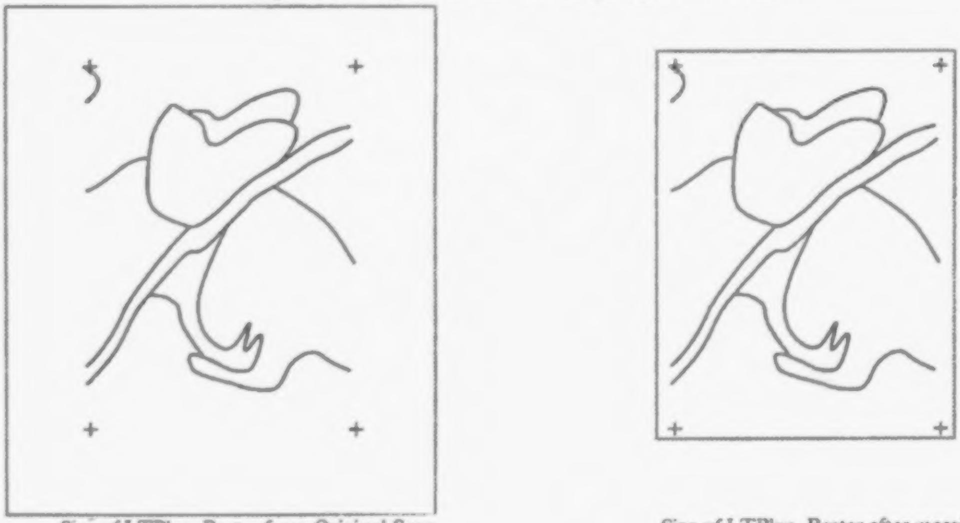
If a margin 20 is used and the pixel size is 20', a 400' wide border remains on all sides of the map that can be used for edgematching. For this project a 400' margin was used and the resulting LTPlus raster files sizes were reduced over 40% (Table 3).

## The effect of resolution on edit and processing times

Following the generalized procedures previously outlined, the next step is editing the raster data in LTPlus. On the first editing pass, low angle intersections are opened and false touches are separated. Lines that do not touch on the manuscript but do touch in the raster are referred to as false touches. False touches and low angle intersections are separated to prevent the creation of erroneous polygons. Then the data is thinned to a single pixel width. On the second pass, if the data is polygon, data gaps are closed and spurious data is deleted.

Separating false touches and low angle intersections on the first pass and repairing identified data gaps on the

**Figure 1** Example of the effect of margining to reduce file size



Size of LTPlus. Raster from Original Scan

Size of LTPlus. Raster after margin

**Table 3.** Results of retaining only the registered map data and a 400' margin

| Scanning Resolution | 75 dpi (2.95 dpmm) | 100 dpi (3.94 dpmm) | 150 dpi (5.90 dpmm) | 200 dpi (7.87 dpmm) |
|---|---|---|---|---|
| LTPlus Raster Size | 0.504 mb | 0.900 mb | 2.034 mb | 3.600 mb |
| Pixel Size | 26.67' | 20.0' | 13.33' | 10.00' |
| *margin* from registered raster (400') | 16 pixels | 20 pixels | 30 pixels | 40 pixels |
| Resulting LTPlus Raster Size | 0.294 mb | 0.512 mb | 1.143 mb | 2.025 mb |
| Reduction in File Size | 42% | 43% | 44% | 44% |

second are done interactively by the user. The number of false touches and data gaps increased at lower resolutions as shown in Table 4. This was expected and it was believed that although there would be more errors the editing time would be offset by faster graphics regeneration after editing zooms. By and large this was true, but not to the extent expected. LTPlus has four pre-defined view levels and at 75 dpi, the views were either too far out to get a good feel for the error or too close to see the data trend. As a result it was necessary to increase the number of zooms and some efficiency was lost.

While the number of false touches and data gaps were less at the higher resolutions editing was slower. This is because more pixels represent a line and more editing decisions can be made. For example, a line one pixel wide at 75 dpi will be three pixels wide at 200 dpi. If the line creates a false touch, turning off one pixel and turning on another to separate the false touch presents no major editing choice to the user. On the other hand at 200 dpi most users begin to "cartographically" edit turning off and on pixels to not only separate the false touch but also create a line that is more visually pleasing.In addition the graphics took more time to regenerate after a zoom, but having to zoom in and out in order to get a proper editing view was unnecessary.

After false touches are separated the data is thinned to a single pixel width in an automated process. The amount of machine time required to thin the data is dependent on the resolution. For the 75 dpi map three passes were required to thin the data to a single pixel width. At 200 dpi, six passes were required. Table 4 compares overall time for completing the editing process at 100, 150, and 200 dpi to 75 dpi.

Once the raster data is edited and thinned it is vectorized and topologically structured. Table 5 compares the time for topological structuring and saving the data. Times are presented as a ratio comparing 100, 150 and 200 dpi to 75 dpi. Again, the higher the resolution the longer the time to create the topology. The time required to save a map was also compared. This is important because it is common practice to save on at least three different occasions in the conversion process. For the 200 dpi map, each save takes over five times longer to complete than the 75 dpi map. This means that completing a map at 200 dpi would take at least 5 times longer if all other factors remained the same.

**Table 4.** Results of the Edit Process

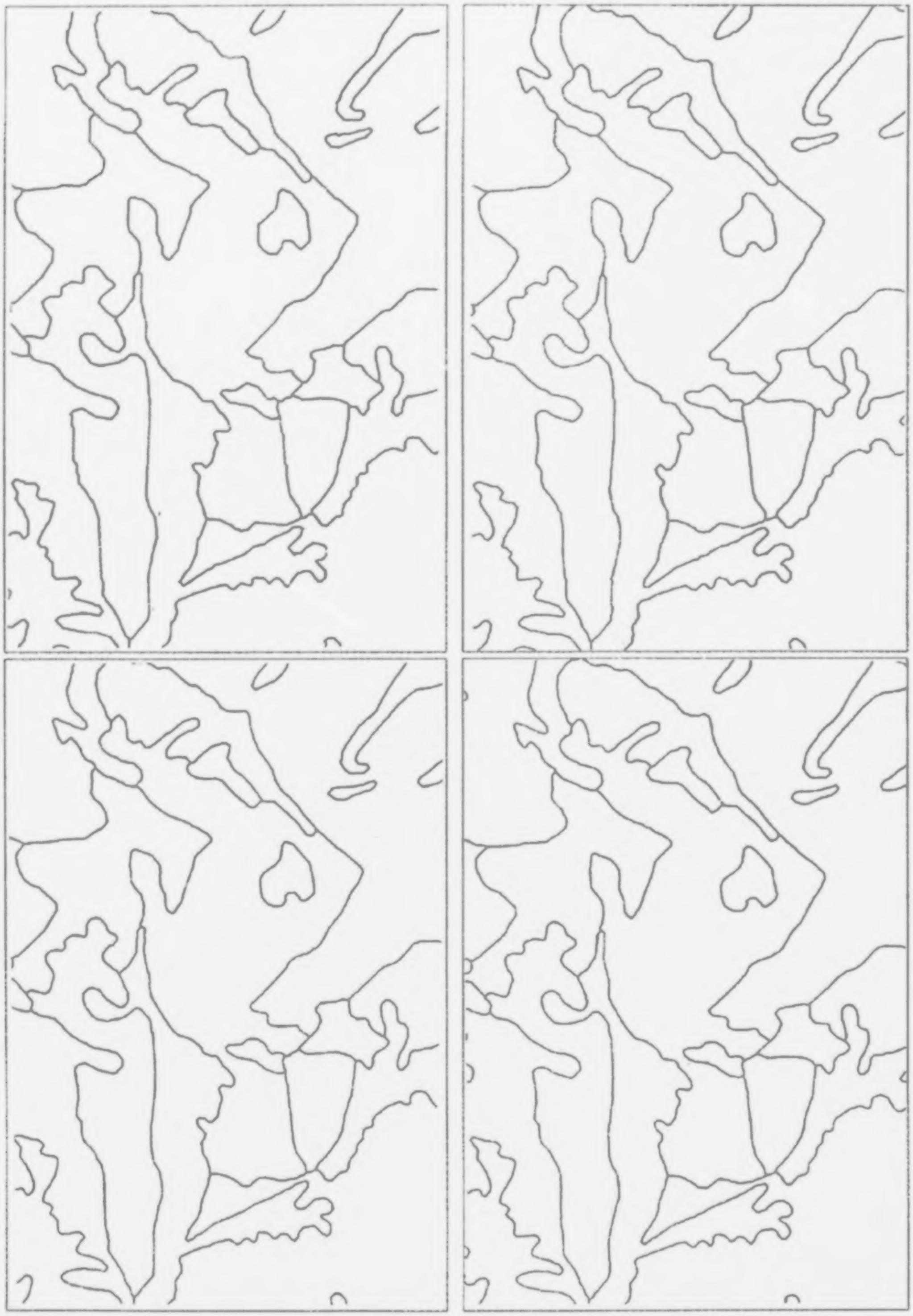
| Scanning Resolution | 75 dpi (2.95 dpmm) | 100 dpi (3.94 dpmm) | 150 dpi (5.90 dpmm) | 200 dpi (7.87 dpmm) |
|---|---|---|---|---|
| Number of false touches | 8 | 8 | 5 | 5 |
| Number of thinning passes | 3 | 4 | 5 | 6 |
| Number of data gaps found | 14 | 10 | 9 | 8 |
| Raster Editing Time, man & machine (Ratio to 75 dpi) | 1 : 1 | 1.3 : 1 | 1.5 : 1 | 2.0 : 1 |

**Table 5.** Results of Topological Structuring and Map Saves

| Scanning Resolution | 75 dpi (2.95 dpmm) | 100 dpi (3.94 dpmm) | 150 dpi (5.90 dpmm) | 200 dpi (7.87 dpmm) |
|---|---|---|---|---|
| Topological Structuring Time (Ratio to 75 dpi) | 1 : 1 | 1.7 : 1 | 2.5 : 1 | 3.9 : 1 |
| Map Save Time (Ratio to 75 dpi) | 1 : 1 | 1.6 : 1 | 3.4 : 1 | 5.4 : 1 |

**Table 6.** Results of Exporting GIS Vector File

| Scanning Resolution | 75 dpi (2.95 dpmm) | 100 dpi (3.94 dpmm) | 150 dpi (5.90 dpmm) | 200 dpi (7.87 dpmm) |
|---|---|---|---|---|
| Output File Size - DLG-3 | 0.377 mb | 0.438 mb | 0.563 mb | 0.703 mb |
| Comparison of DLG-3 File Size (Ratio to 75 dpi) | 1 : 1 | 1.2 : 1 | 1.5 : 1 | 1.9 : 1 |



## The effect of resolution on vector output

The vector data was exported in an ASCII format, USGS Digital Line Graph-3 (DLG-3). USGS DLG-3 file is composed of three separate but related elements nodes, lines, and areas. Table 6 shows the size of the DLG-3 files and compares the 75 dpi file to the higher resolution files. Predictably, the size of the final vector output file increases with resolution. This relationship will continue once the data is imported into a GIS and increased analytical processing and graphics times generation can be expected.

The vector data is displayed in Figures 2-1 through 2-4. These figures show a section of the map plotted at the original drafted scale of 124,000. To the naked eye at the original map scale, the vectors appear similar. It is possible to see the effects of the resolution at tight corners and low angling lines. Also when the data is presented at larger scales the smoothing differences become more pronounced. Still, all four plots maintain the on-the-ground accuracy of the data. When the plots are compared with the original drafted data, the 100, 150, and 200 dpi resolution maps consistently lie within one line width of the original data. The 75 dpi map achieves this on over 85% of the map. See Figure 2-1 through 2-4.

## Conclusion

Scanning resolution impacts all phases of the data conversion process and carries into the GIS analysis process. The project attempted to evaluate the general effect of scanning resolution on LTPlus data conversion by eliminating activities that could have minor influence at certain resolutions. Varying scanning thresholds and thinning vertices could have optimized the higher resolution results.

For this project, machine processes not manual editing were most impacted by changes in resolution. Primary machine processes include importing a scan file, thinning raster data to a single pixel width, vectorizing, creating topology, thinning vertices, exporting the vector data, and saving. Carrying a map through from import to export may require executing a machine process multiple times. For example, the user may save after redefining the map margin, before deleting all remaining spurs on a polygon map, and at the end of the session. Or the user may find an error after vectorizing and creating topology which can only be repaired by editing the raster map and vectorizing and creating topology again. Saving three times and assembling vectors twice would take over 4.5 times longer at 200 dpi than at 75 dpi. The cumulative effects of these repetitive processes should not be underestimated when selecting a scan resolution.

For this project editing the 200 dpi map took twice as long as the 75 dpi map while creating topology and saving took over 4.5 times as long. A key reason why the resolution impacts to manual editing time were so low was because the data was scan ready. Scanning reproduces the original manuscript. If a map is well manuscripted, lower resolution scans will have a few more false touches and data gaps than higher resolutions scans. For this project, the 75 dpi map had 8 false touches and 14 data gaps while the 200 dpi map had 5 false touches and 8 data gaps. The total number of number of errors that were the result of manuscripting and/or scanning was less than 2% of the total number of data segments.

File sizes cannot be overlooked, as resolution increases so does file size. This is true throughout the conversion process and knowing how the software handles the raster and associated files is important. When a map is acquired in a program like LTPlus, LTPlus makes a duplicate of all associated files (21). It also copies a number of files into RAM including the raster image, vector image, edgematching vectors, registration, and attribute information. While it is highly advantageous to compress files when they are not in use, the raster map and its associated files will not be compressed during conversion.

Finally, all resolutions produced data that fell within the positional on-the-ground accuracy of the original map. There is a visual difference between resolutions with the 200 dpi map producing the most flowing linework. The 75 dpi map produced the most raster-like linework with 100 and 150 dpi maps falling somewhere between. It is the GIS users decision as to whether the vector ends justify the resolution means.

## Appendix A

**LTPlus system requirements**

Minimum LTPlus requirements

- Hardware
  - PC
    - 80386 (or better)
    - Hard Disk _300 Mb
    - Memory _8 Mb
    - 80387 Math CoProcessor
    - BIOS & register-compatible
    - VGA graphics card
    - One parallel port, two serial ports
    - Floppy disk &/or tape drive
  - Monitor/Graphics
    - Color multisync
    - (400x600 resolution)
  - Mouse for scan editing
    - 3 button, serial
  - Scanner
    - Producing standard scan formats (e.g. .RLC, )
  - or
  - Digitizer
    - Varied
  - Plotter
    - Varied (HPGL)
- Software
  - Unix or Xenix Operating System
  - Text Editor (vi)
  - LTPlus

Infotec Analysis Components

- Hardware
  - PC
    - Compaq 386/20e
    - Hard Disk - 320 Mb
    - (Xenix Partition = 288 Mb)
    - (DOS Partition = 32 Mb)
    - Memory 10 Mb
    - 80387 Math CoProcessor
    - BIOS & register-compatible
    - VGA graphics card
    - 1 parallel port, 8 serial ports
    - Floppy disk & tape drive
  - Monitor/Graphics
    - NEC 20" multisync color
    - (1024x768 addressable)
  - Mouse for scan editing
    - Logitec 3 button, serial
  - Scanner
    - Houston Instruments LDS 4000
    - (Monochrome)
  - Plotter
    - Hewlett Packard 7585B
- Software
  - Xenix Operating System v. 2.3.2
  - Text Editor (vi)
  - LTPlus version 2.30
  - MS-DOS version 3.2 (for scanning)

GIS'92

Symposium

Vancouver
British Columbia
February 1992

# Scanning technology: considerations in the preparation of scanner-ready manuscripts

**Scott Oerding,** *GIS Specialist*
**Daniel Haskell,** *Technical Leader*

Infotec Development Inc., GIS/LIS Facility
500 N.E. Multnomah, Suite 329
Portland, Oregon. 97232 USA
Tel: 503-238-0034
Fax: 503-238-1224

## Abstract

*With the increased availability of scanning technology, more organizations are recognizing it to be a "quick and easy" means of GIS data capture. However the speed and accuracy of this approach is strongly tied to the quality of the source manuscripts. Not all manuscripts are scanner ready. This paper first examines the considerations necessary for the creation of scanner ready source maps. Secondly, it discusses criteria for evaluating the suitability of existing manuscripts for this type of capture.*

## Introduction

Scanning technology provides a high tech solution to the problem of GIS data capture. In a short period of time a manuscript can be run through a scanner, loaded in the editing software, cleaned up, structured topologically, attributed and added to a database. In fact, given that much of this processing can be performed by the computer with minimal user input it could be said that scanning is also a "smart" technology. Unfortunately its speed and high tech smarts depend heavily on a low tech issue: manuscript quality. Poor quality or inappropriately drafted manuscripts can negate all of its advantages. To receive the full benefit of scanning technology it is necessary to start with scanner ready manuscripts.

What makes a manuscript scanner ready? In the last 14 months Infotec's GIS/LIS Center has used scanning to capture over two thousand maps for the U.S. Forest Service and, based on that experience, would like to offer a few suggestions. These suggestions have been culled from our experiences scanning with a Houston Instruments LDS4000 Plus scanner and editing with the Line Trace Plus (LTPlus) software.

## Defining the problem

Poor quality or inappropriately manuscripted maps require much more time to capture. In many cases the capture time can even equal what could be achieved through manual digitizing. This happens because scans of poor manuscripts require a large amount of manual editing before they can be processed: A human must go through the scan, separating and/or completing the object images to make them discernable for the computer. While scanning is usually computer intensive, poor manuscripts make the process labor intensive. The goal in creating scanner ready manuscripts, then, is to draft images that can be processed with as little manual editing as possible.

There are two occasions when it is important to consider scanner readiness: (1) when preparing new maps and (2) when evaluating pre-existing maps for capture. The first section of this paper addresses how to draft scanner ready manuscripts. The second section addresses how to evaluate existing manuscripts.

## Drafting a solution

The best way to ensure that your manuscripts are scanner ready is to draft them that way. To do this it is necessary to "think digital". This is accomplished by imagining how the manuscript will look to the scanner. The following is a list of areas of concern in creating a manuscript that will be easily scannable:

**Stable base media** Most GIS professionals know that paper and velum are not appropriate media for manuscripts. When the humidity and temperature in the room changes so does the size of paper or velum maps in that room. They are also more vulnerable to stretching and tearing. However, when manuscripting for a one time use, like GIS data capture, it is tempting to cut corners. Don't. Mylar may be more expensive, but it will help safeguard the accuracy of the data. An opaque, single matte mylar works very well.

**Ink vs pencil** Pencil usually works rather poorly for scanning and should be avoided. Penciled lines often smear, creating an undesirably dark background which could obscure the linework. Also, even though penciled lines look dark to the human eye, the scanner often "sees" them as inconsistently opaque causing gaps and splotches. A bold black pen seems to be the best choice.

**Linework** It is important that linework be drawn as evenly and smoothly as possible. Overly thick sections may clump together in the scan, obscuring the linework. Overly thin lines may be lost or only partially captured. Dashed lines are scannable but not recommended as they often require additional editing time to connect. Smears and smudges on the media can make sections of the linework look like a muddy mess to the scanner. In all it is best to use a pen that produces an opaque black line and try to maintain a uniform level of pressure while drafting. The size of the pen point would vary depending on the estimated accuracy of the data.

**Registration points** It is important the map contain four accurately placed registration points, preferably outside the data. If these points are not placed in their correct geodetic locations the data will be distorted. The further off the placement the worse the distortion. Ideally each map would have four simple "plus sign" tics inked onto each corner of the map. Stick-on tics are not recommended because they can be moved by the traction wheels of the scanner.

**Colored lines** When using a monochrome scanner, it is best if all the linework be drafted with the same color pen. Almost all colors of ink will scan, but some scan better than others. When multiple colors are used it may be impossible to make the scan light enough to capture one color without making the others indiscernible. Yellow comes the closest to being non-scannable and works well for lines which are included for reference only (such as the perimeter of a quadrangle).

**Object attributes on a separate sheet** It works best to draft attributes onto a separate sheet. Attributes included on the manuscripts will require editing time to remove. The separate sheet could be a registered clear acetate overlay or a blueline copy of the manuscript. The attributes should be drafted in a dark ink for maximum legibility.

**Map name/number** Whenever possible draft the map name/number somewhere within the tics. This allows the worker to verify the identity of a scan file at edit time.

**Gaps in linework** The elimination of gaps in the linework of both polygon and line data is important. All lines should extend to meet boundaries and/or the map's geodetic or planar edge. This practice can eliminate the need for interpretation of the data by workers during the editing process and, thereby, speed editing time.

**Edgematch all mylars** During the manuscripting process carefully edgematch each mylar with its neighbor. Taking the time to ensure that the data matches across each map sheet will greatly reduce the need for interpretation, editing and re-work at the time of capture. Adding the name or number of each edgemap to the corresponding edge of the manuscript will speed the edgematching process as well as make future updates easier.

**One theme per manuscript** Composite themes generally do not lend themselves well to digital data conversion through scanning. They tend to be very complex, confusing and usually require a great deal of time to interpret and edit. This is especially true if the manuscripts were drafted with a single color pen and/or they are being scanned with a monochrome scanner. Each theme must be carefully separated from other background information. These maps also require more processing time and storage capacity.

## Evaluating pre-existing manuscripts

Unfortunately, there are many organizations that have large catalogs of pre-existing manuscripts they need converted to a GIS digital format. More often than not these manuscripts were not drafted with GIS, let alone scanning in mind. While the ideal solution might be to custom manuscript data explicitly for scanning, it may not always be feasible. In these cases it is important to be able to determine whether or not scanning is feasible at all.

When evaluating pre-existing manuscripts for scanning it is useful to use the following approach:

(1) Evaluate the maps against the suggestions listed in the "Drafting a solution" section (see appendix A for a checklist)

(2) Use the suggestion list to determine ways in which the maps could be made more scanner ready.

(3) Use the list in the "Handling Special Cases" section (below) to determine if your maps may be captured using a special procedure.

(4) Beta test the capture process with a representative sample of the maps. Determine whether the time and resources that are required will be acceptable. Consider both human and computer resources.

(5) Consider re-manuscripting. If it appears that capturing your existing maps would require more time or resources than are acceptable, it may be preferable to trace the maps onto new sheets and scan the new sheets. In practice we have found it to be more cost effective to trace and scan than to hand digitize. This may sound outrageous, but digitizing is a slow tedious and inaccurate process compared to scanning.

## Handling special cases

Below is a list of situations that can make otherwise perfect manuscripts un-scannable or poor manuscripts scannable. It addresses a few of the special cases that can make or break the chances of capture through scanning.

**Multi-color manuscripts** Multi-color manuscripts can be readily scanned IF you have a color scanner. Some color scanners come bundled with color image processing software which allows users to improve the image before taking it into the GIS capture software.

**Multi-color composites** When the manuscript contains a number of themes which are coded by color linework, it may be possible to separate the themes using a color scanner. If the roads, for example, are in red and the rivers are in blue the two sets of linework could be separated through use of the image processing software bundled with the scanner.

**Smudged, smeared, or dirty manuscripts** Pay careful attention to the condition of the manuscripts. An excessively smudged, smeared or dirty map may be useless for scanning. It may be possible to salvage such maps using variable thresholding on a monochrome scanner. A better option might be to perform some editing on a grey scale or color image of the map before taking it in to the GIS capture software.

**Paper and pencil** From an accuracy standpoint paper maps and pencil linework are horrendous. We have seen variation in paper manuscripts of greater than a quarter inch (which equates to a greater than 500 foot discrepancy on a 1:24000 scale theme). As for pencil, it smears worse every time it goes through the scanner. The more the linework smears the less accurate it becomes. The smearing may eventually make the background so dark it obscures the data.

**Monochrome composites** If the manuscripts in question are monochrome composites consider either digitizing or tracing the data on to a separate sheet. Such maps are labor intensive in the extreme.

**Orthophoto-quads and aerial photos** Some organizations have drawn data directly onto orthophoto-quads or other types of aerial photos. Orthophoto-quads have dark backgrounds and aerial photos have cluttered backgrounds that make them unsuitable candidates for scanning.

## Conclusion

Scanning of geographic data for use in a GIS can be a quick easy process where the computer does most of the work; or it can be a slow and tedious one, requiring extensive human input. The nature of the experience hangs on manuscript quality. Properly manuscripted maps allow the user to take full advantage of their computer's intelligence. Poorly manuscripted maps require much user time and input.

## Appendix A

A brief checklist of points to consider when preparing manuscripts or evaluating existing manuscripts for scanner readiness:

____ Stable base media

____ Ink linework

____ Consistent linework

____ Four accurately place registration points

____ Single colored lines

____ Object attributes on a separate sheet

____ Map name/number included inside grid tics

____ No gaps in linework

____ Data edgematched with adjacent maps

____ One theme per manuscript

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Data collection for GIS applications, using scanning in conjunction with automated raster to vector conversion technology.

**Jeff W. Marceil** (presenting author)

Scan Conversion Services inc.
4228 Manor Street
Burnaby, BC
V5G 1B2
Tel: (604) 439-0054
Fax: (604) 439-1459

## Abstract

*With the 'coming of age' of scanning technology, considerable efficiencies in data collection for GIS applications can now be realized. Perhaps even more significant than the scanning component of the process (although the quality and accuracy of a scanned image is critical) are the developments in raster to vector conversion technology. This technology utilizes the end product of a scanner (raster or bit map data) and produces an intelligent vector based file. It can then be cleaned and built to produce a fully functional, accurate GIS product.*

*Proper utilization of the automatic capabilities, found in both raster to vector and GIS software, can result in the conversion of source maps to fully structured GIS files with very little operator time required. Consequently, the cost of data collection can be substantially less than other methods of conversion. To ensure the most cost effective execution of such a project, it is imperative that several critical steps be followed to maintain conversion efficiencies.*

## Introduction

In recent years it would appear that the level of knowledge obtained by many digital graphic users on the subject of scanning has increased substantially. This is likely due to their exposure to the technology either directly or through the increased amount of literature and articles available on the subject. As a member of a scanning service bureau, namely, SCAN Conversion Services Inc., the writer respectfully makes the following observation. Many digital mapping users, feel they already know all they need to, on the subject of scanning. However, many still have some misconceptions about the technology, and its capabilities. Case in point: often managers of a conversion project have been heard to say, "But, with scanning, I'm going to end up with everything on one layer (level, feature class, coverage, etc...)" or, "all the line work in a scanned file is made up of small broken line segments...", or "but, a scanned file is in a bit map format, what can I do with that?" If approached correctly, the end product produced through a scanning system should satisfy the end user, both cartographically and functionally.

With the objective of clearing up these misconceptions in mind, this paper is not intended to impress the reader with a lot of technical jargon, but to examine the issues of scanning. Hopefully, some constructive suggestions can also be made, to help the technology produce it's full capabilities in data collection.

## Raster to vector technology

The principal by which most scanning systems operate is fairly simple in concept.

1) Using a high resolution scanner (available in all shapes and sizes) which can physically accommodate the desired map, the document is optically read and that image is converted to a bit map or raster image. This raster image is much like a photocopy in a digital format.
2) The raster image is then converted to a vector image. This is accomplished in various different ways depending upon what system is used. Some systems require considerable interactive work by an operator, such as, on screen raster tracing/ digitizing. Other systems have automatic vectorization, in which case, the raster to vector conversion is done with little or no operator intervention. Such systems utilize artificial intelligence and parallel processing to perform the vectorization. Yet, other systems incorporate techniques from both of the above procedures to carry out the conversion.
3) In systems with automatic vectorization, the vector file is then cleaned up interactively by an operator, using highly specialized software. Any areas in the map which the vectorization system could not resolve, are repaired. This produces a cartographically correct digital image.
4) Using a compatible exchange format, the digital file can then be, imported, transformed, cleaned, built, edgematched and further processed as required, in a GIS environment. These functions can also be carried out using utility software, working in conjunction with CAD or Digital Mapping software.

## Project analysis for optimum scanning results

When approaching a digital mapping project for AM/FM or GIS applications, there is one question that certainly now enters the minds of most project managers. "Is scanning a feasible method to accomplish the desired end result?" The answer to this question is a resounding, "yes!" In this regard, there are some key questions which must be considered, when evaluating the place of scanning technology, in a conversion project.

Perhaps the most important aspect, of evaluating a conversion project using scanning technology, is to approach the exercise with an open mind. Of course there will be specific goals or objectives that are desired upon completion of the project. However, be reasonably flexible regarding new methods to reach that end.

Rumours you may have heard might also sway your opinion on scanning. A colleague's experience with scanning, may be an isolated incident. It could be the result of several variables which were not completely explored. Consequently, the end product was not completely satisfactory. Of course this can happen regardless of what conversion process is used.

As with any situation, when dealing with scanning as a means of conversion, there are always many options open to the user. Specifically, with regards to how a desired end product will be accomplished.

## Accuracy of scanning

The first question a digital map user must ask him/ herself, is regarding the purpose of the data, and how accurate it needs to be. More specifically, will the data be representational or positional. Also, will the data be related to an existing base map or will it be the base map itself? We all want to have "perfect accuracy" in our mapping. However, what usually brings us back to reality, is the cost to achieve this kind of accuracy. Especially, when quotes come in and budget day comes around. In other words, the more accurate the desired end product, the larger the price tag. Perhaps this is scanning's biggest advantage. If handled properly, scanning can accommodate the accuracy issues while coming in at a reasonable cost.

There is one key factor to remember when dealing with scanning technology. Scanning will do a wonderful job of reproducing a hard copy map with precisely the same accuracy that was present in that original document.

Many times digital mapping users are convinced that their source material is perfectly accurate. This is until it has been scanned and errors reveal themselves. For instance, when a mylar plot of the converted file is produced, it lines up perfectly with the original map. However, when distances are measured within the file, they do not correspond perfectly with the dimensional annotation in the original. Usually those distances are very close... but not perfect. If the source material was produced using manual systems incorporating human intervention this error can only be expected. Consequently, without additional processing, scanning technology cannot be expected to produce a file more accurate than the original.

## Augmenting the accuracy of scanned files

As stated earlier, there are always several options opened to the user, to achieve a desired result. One technique which can greatly increase the accuracy of a scanned file, is to utilize the "Rubber Sheeting" or transformation capabilities found in most GIS systems. As mentioned earlier, this type of capability is also available from several vendors as a separate utility program to run in conjunction with existing application software.

The procedure to increase the accuracy in a map is, to tie "control", such as UTM or Latitude and Longitude coordinates to geographic features in a digital file. Upon inputting these coordinates into the transformation software, it will massage the map relative to the supplied coordinates. If a tic or grid system is used as control, those features will be moved to their respective, precise coordinates. All other detail in the file is then transformed relative to that grid or tic system. Many transformation software packages can calculate an unlimited number of control points. The more control points used in a map sheet, the better the result.

In the final analysis, prior to using any special techniques and software to resolve accuracy issues, the determining factor of a successful scanning project, is the quality and accuracy of the source material that was used.

## Source document preparation

The preparation of source maps plays an important role in determining what level of success may be achieved in a scanning project. This preparation can facilitate the best production efficiencies, resulting in cost benefits for the overall project.

The first rule in scanning is, not to try and shove something down the scanners throat, that it doesn't like. The first source material chosen for a project, may not be suitable for scanning conversion, in its initial state. This may be due to the quality of the original, or how the features in the map where compiled. Some streamlining of the system used in source preparation may be needed. Or, by using a different source map, which lends itself better to scanning, the conversion costs may be affected substantially.

The objective of source preparation, is to result in material which allows for the full utilization of the automatic vectorization features in the scanning process. The efficient conversion of entities in a map, decreases the need for operator time. Such time, is required to clean up areas of the map, which the automatic vectorization system left unresolved.

The following are general guidelines, for choosing scanning source material. It is best to look for maps processing these characteristics.

a. Clean line work — The less background noise, or fuzziness of linework present in a source map, the easier it is for the scanner to interpret the information. Mylar or some other stable media is preferred to produce the most accurate results.

b. Intersecting lines — The more lines that intersect in a map, the higher the chance that the automatic vectorization may break the line or continue along the wrong line. If this is widespread, more clean up will be required, which affects the price of the conversion. One method which has been found to be effective, is by separating certain features onto different source maps. i.e.) Mylar Separations

1. Forest polygons
2. Planimetry
3. Hydrography
4. Contours

Separating features in this way, eliminates the number of intersecting lines on a map when a compilation of features are required .

Energy Mines and Resources, Mapping and Remote Sensing Sector in Ottawa has utilized this in the scanning of 1:250 000 and 1:50 000 NTS maps with considerable cost benefits.

c. Line types — A clean solid line is the preference of most scanners. Generally, any dashed or dotted line will either require tracing or repair. The clean up software used in conjunction with scanning can quickly repair these lines to a continuous solid line very efficiently. The end product will always be excellent (depending upon the operator), but the price will be affected somewhat.

## Pilot projects

Enough cannot be said for the importance of pilot projects. There are many areas where a conversion project manager, or the subcontractor could misinterpret the requirements of the project. Of course, this holds true no matter what the project, or method of conversion.

Carrying out a small test, or pilot project, allows the scope and details pertaining to the total project, to become much more defined. The methodology to be followed for the duration of the project, can be streamlined and set in place. A pilot project simply allows for some good constructive communication at all levels of the project hierarchy. It can also determine, if in fact a contractor has the resources or expertise to effectively carry out the total project.

## Meshing of scanning and GIS technologies

Certainly scanning has come along way, and the capabilities of the technology are far beyond what they were just a few years ago. However, perhaps equally as important, is the evolution in understanding of the technology by it's users.

## Case study

A fairly important realization at SCAN Conversion Services Inc. has been brought to light. It was determined that we were cleaning up scanned files too much for GIS applications. We had always been very careful to produce cartographically correct digital maps as an end product. Consequently, we had cleaned up all vector files until they were a thing of beauty.

While thinking through a proposal for a client, regarding the conversion of several timber and soil type maps, we asked ourselves the following. The client wanted the end result to be in ArcINFO format. Therefore, did we need to clean up the vector file so meticulously if it was to be topologically cleaned later? In other words, could we leave breaks in lines, and intersections which the scanning system left unresolved, if the cleaning function in ArcINFO could do it for us?

We ran some tests, and determined that this was in fact the case. By setting both the dangle, and fuzzy tolerances at appropriate settings, the errors left purposely in the file where repaired automatically. This resulted, in very little time spent cleaning up the files at the scanning level. The same amount of time was spent at the GIS level of the conversion process. This is because, that cleaning function would have needed to be undertaken in the GIS, regardless.

The result of this process refinement, has been a savings in time and money. This has been passed onto that client, and all of our other GIS clients. Using this method of conversion, SCAN Conversion Services Inc. has completed numerous map sheets for B.C. Ministry of Forests, Re-Inventory Contractors. Both cost effectiveness and a high level of quality have been found to be the results.

## Conclusion

Scanning has proven itself to be an exiting and effective form of AM/FM and GIS data collection. As apposed to being an alternative form of conversion, as it has been perceived in the past, scanning is now becoming the preferred method of conversion for many applications. This is due to the considerable time and cost savings that can be realized from using scanning. Highly accurate digital mapping can be achieved, if scanning technology is utilized properly.

It is also worth noting that, at last years GIS'91 conference, there were only a few papers on scanning technology. For the most part they were lumped in with other unrelated presentations. This year GIS'92 feels the technology is important enough to represent it's own exclusive scanning forum. It looks like scanning is here to stay.

# D7 Integrating images with GIS

---

**Utilizing satellite data and GIS to map land cover change**

Paul Maus et al., Pacific Meridian Resources, Emeryville, CA, USA

**Replacing aerial photo interpretation in resource inventories with integrated digital images and GIS**

Frank Hegyi et al., Ferihill Technologies Ltd., Victoria, BC, Canada

**Managing forest resources with raster tools**

Andrew Bury, ERDAS, Inc., Atlanta, GA, USA

**Using ancillary data in post-classification modelling to increase the accuracy of conifer species classifications from LANDSATÔ data**

Mike Golden et al., Pacific Meridian Resources, Portland, OR, USA

---

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Utilizing satellite data and GIS to map land cover change

**Paul Maus**
**Vaughan Landrum**
**Jan Johnson**

Pacific Meridian Resources,
In Residence at Nationwide Forestry Applications Program,
USDA Forest Service, Salt Lake City, UT 84116

**Henry Lachowski**

Nationwide Forestry Applications Program,
USDA Forest Service, Salt Lake City, UT 84116

**Bruce Platt**
**Mike Schanta**

Mark Twain National Forest,
USDA Forest Service, Rolla MO

## Abstract

*Monitoring vegetation changes in the landscape over time is an important concern that land use managers have been confronted with for many years. This problem has increased in importance as the land base available for natural resource extraction has decreased and the demand for the resources has increased. Landsat Thematic Mapper (TM) digital imagery has been available since 1982 and continues to be an excellent source of information about the Earth's land cover. Two TM images were acquired, seven years apart, over an area of land managed by the Mark Twain National Forest in south-central Missouri. Band 7's from both images were combined and an unsupervised classification was run to classify five categories of change in the landscape. The classification was integrated into a GIS database which contained land ownership and management unit boundaries. A prototype map and table were produced for selected management units to determine how well the Forest Service is meeting the objectives of the Forest Plan.*

## Introduction

The use of remote sensing systems for mapping land cover change has been well documented and continues to be one of the most promising emerging technologies (Colwell et al., 1980, 1981; Nelson, 1982; Lee, 1980). Satellite imagery has many advantages over other remote sensing systems for mapping land cover change. The benefits, including repeatable coverage, high spectral and spatial resolution, and highly accurate image registration methods have helped make satellite imagery an excellent vehicle for mapping land cover change. These advantages are augmented by the fact that satellite data exists in digital form, making it readily available for integration into a geographic information system (GIS).

Early problems associated with mapping land cover change from digital satellite imagery have included poor registration, lack of established change detection procedures, and poor methods for determining the boundaries and acreages involved in change (Malila, 1980). Improvements have been made in all of these areas due to advances in computer hardware and software, combined with new methodologies for manipulating digital data, including classification and positional accuracy calculation procedures.

## Background and study area

This paper summarizes the cooperative project conducted by the Mark Twain National Forest and the Nationwide Forestry Applications Program (NFAP) under the "Integration of Remote Sensing" (IRS) program. Contract support to IRS is provided by Pacific Meridian Resources of Emeryville CA. Analysis was completed on a SUN workstation with ERDAS and ARC/INFO software.

The Forest Land and Resource Management Plan for the Mark Twain National Forest was begun in 1980 and approved in 1986. The Plan's mid-point review was begun in 1991. The project developed from a need to gather and anlalyze information to monitor Plan accomplishments in conjunction with the mid-point review. The Forest Plan sets standards for vegetaion change by management prescription area in which data is needed for private and federal lands. The project integrates the use of satellite image analysis in conjunction with a GIS to determine the type and amount of vegetation change during the Plan period.

The Mark Twain National Forest is located in the area of southern Missouri known as the Ozarks. The Forest encompasses an area about 235 miles east to west and 100 miles north to south. There are nine districts located on nine separate units. Net ownership of National Forest System lands in 1990 was approximately 1.5 million acres, an average of about 50% within the proclaimed National Forest boundary. Ownership within the individual units ranges from about 30 percent to 70 percent. The Forest is located in the transition zone of several major ecological plant associations. The oak-hickory, oak-hickory-pine, cedar glades, southern floodplain forest and bluestem prairie all converge within the Forest.

The project described in this paper was conducted on one of the nine units on the Forest called the Fristoe Unit. Figure 2 shows the boundaries of the Fristoe Unit and the three shaded management unit boundaries used in the sample analysis described in this paper.

## Analysis methods

### Image classification and polygon creation

Landsat TM data from August 1982 and August 1989 was acquired for a portion of the Forest. The August 1982 image was registered to the August 1989 image which had been precision geocoded and terrain corrected. The image registration was tested by performing a visual inspection of individual bands from the different dates in separate color guns on an RGB (red,green,blue) display. Image registration of the two scenes was excellent because of the careful preprocessing procedures used.

Clouded areas create problems for change detection analysis and will classify incorrectly as land cover changes if not removed. Clouds were confined to a small portion of the images but were most persistant in the southern third of the project area. A "cloud mask" was created from an initial unsupervised classification and areas classified as clouds were removed. This method creates holes in the data but prevents misclassifications and results in fewer errors in the analysis.

Landsat TM data has seven spectral bands or channels in which it gathers information about earth surface features. In this project only band seven's were used from each of the images. Band seven's from each of the masked images were combined into a single file containing the two bands. Band seven was used because the digital values in this portion of the spectrum (middle infrared) were consistent for earth features that showed little or no change between the two image dates.

An unsupervised classification of the combined band seven image was completed with 35 classes. The unsupervised clustering program and classifier separated out various degrees of change between the two bands and created an output classification with 35 information classes. The classification was further aggregated into five categories which represented various classes of land cover change. The land cover change classes are shown in Figure 1.

The evaluation and labeling of the unsupervised classes was based on visual and statistical analysis and was guided by the expertise of Forest Service personnel who were familiar with the forest. A forest history database, containing a history of activity in individual stands, was also available and proved valuable for assigning the unsupervised classes to land cover change categories. A ground-based accuracy assessment was initially completed for a portion of the project area and is currently being completed for the entire project area.

The resulting land cover change map was subjected to a series of scanning filters which served to create homogeneous pixel areas of land cover change. An assortment of majority scans were used to firm boundaries between classes and eliminate outlier pixels in the classification. Vector polygons were created from the scanned classification and polygons less than two acres were eliminatedfrom the analysis. Additional vector GIS layers of land ownership and forest management unit boundaries were created from the Forest Service's Cartographic Feature Files (CFF's).

**Figure 1. Land cover change classes derived from the Landsat TM classification.**

| Land Cover Change Class | Description |
|---|---|
| 1. Open Areas | Areas that were classed as open in 1982 and in 1989. |
| 2. Change | Areas that have not had recognizable change in forest canopy density from 1982 to 1989. |
| 3. Regrowth Areas | Areas that show an increase in woody vegetation since 1982. These would include recent and older clearcuts. |
| 4. Moderate Change | Areas that show a moderate decrease in forest canopy density since 1982. These include commercial thinnings, moderate fire killings, etc. |
| 5. Major Change | Areas that show a major decrease in forest canopy density since 1982. These would include clearcuts, conversion of forest to pasture, fire kill, etc. |

## Results and discussion

The preliminary results of the Landsat TM classification were excellent and the methods used proved to be easily understandable and readily applied to large areas of land. An initial accuracy assessment has been completed for a single 7.5 minute quadrangle and further assessment is underway for the entire Fristoe Unit (approximately twenty, 7.5 minute quadrangles). A random sample of land cover change polygons was selected and checked on the ground and with the Forest's activity database. The preliminary accuracy assessment demonstrated that 46 out of 51 sites were classified correctly for an overall accuracy of 90%.

### Forest plan monitoring needs

An important part of Forest Plan monitoring is measuring the progress toward attaining the Desired Future Condition (DFC) for each of the management prescriptions and areas. Each management prescription has a specific goal, or limitiation, for the amount of permitted vegetation manipulation. Questions that may be answered using this technology are: Where and what types of change are occuring? How much change is occuring? Which has more impact — Forest Service activities or private activities? Is the Forest following Standards and Guidelines? Achieving it's goals?

Comparision of the amount of change or disturbance during the period versus the amount desired under the Plan is one measure of how well the Forest is progressing toward attaining the desired vegetation conditions. In addition, the Forest is charged with monitoring the effects of these changes on adjacent lands as well as the effects upon National Forest lands by other agency and private land activities.

Monitoring of changes to vegetation was tracked using the process developed in this project. Even-aged management areas show clearcuts as major change and thinnings as moderate change. The shapes of the polygons are variable in size and shape. Size limits of the clearcuts can easily be determined and compared to the maximum size allowed. Total acres in the 0 to 9 age class and percent of the area, can be determined and compared to the goals of that management unit. Distribution (grouped versus scattered) of these activities is also very important and can be monitored with the methods developed here.

### Need for resource information

Basic resource data for the Forest is obtained from a detailed field inventory. It is updated based on management activities and new inventories and forms the basis for district management activity planning and scheduling. Data is only available for National Forest lands. Approximately 3 to 7 percent of the Forest is reinventoried each year.

With government ownership ranging from 30 to 70 percent on the units, data is often needed on private as well as the National Forest lands to assess total impacts. The Mark Twain National Forest received over 90 appeals of proposed activities in the past two years. With it's proximity to large population centers and the large number of people living within the Forest boundaries, it is inevitable that people will disagree with some management decisions. There is a real need to show these proposals and their effects in a visual format (e.g. images and maps) in order to make easily understandable presentations to land managers and the public. Analysis has to be on larger areas and on a landscape basis rather than simply the affected acres and immediate surroundings.

The objective of this project was to apply change detection procedures in monitoring Forest Plan. Results of the imageanalysis are in the form of a digital data layer of "change". This layer, in addition to the other data

**TABLE 1. SAMPLE PRIMARY ANALYSIS:**
The following table is an example of the primary analysis product of this project. It was created by overlaying a GIS layer containing the boundaries of three selected management units over the change detection GIS, and producing a summary of change by prescription and land status (public vs. private). This type of summary can be generated by prescription, by management unit, by ranger district, etc., limited only by the layers of data input into the Forest GIS:

| MANAGEMENT PRESCRIPTION | LAND STATUS | 1982 to 1989 CHANGE CLASSIFICATIONS | | | | | | | | | | | TOTALS | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NO CHANGE | | AG/OPEN | | REGROWTH | | MOD. CHANGE | | MAJOR CHANGE | | | | |
| | | Acres | % | Acres | % | Acres | % | Acres | % | Acres | % | Av. Size | Acres | % |
| 3.1: Natural | NF | 3055 | 88 | 30 | 1 | 153 | 4 | 38 | 1 | 195 | 6 | 7.0 ac. | 3472 | 68 |
| Vegetation | PVT | 931 | 57 | 256 | 16 | 216 | 13 | 18 | 1 | 224 | 14 | 11.8 | 1645 | 32 |
| Communities | Total | 3986 | 78 | 286 | 6 | 369 | 7 | 56 | 1 | 419 | 8 | | 5117 | |
| 3.4: Wildlife | NF | 5870 | 84 | 164 | 2 | 707 | 10 | 31 | .4 | 256 | 4 | 8.5 | 7028 | 67 |
| Habitat | PVT | 1445 | 42 | 948 | 28 | 781 | 23 | 86 | 3 | 177 | 5 | 4.3 | 3437 | 33 |
| Emphasis | Total | 7315 | 70 | 1112 | 11 | 1488 | 14 | 117 | 1 | 433 | 4 | | 10465 | |
| 6.1: Semi-Primitive | NF | 2889 | 90 | 43 | 1 | 258 | 8 | 25 | 1 | 6 | 0 | .6 | 3222 | 93 |
| Non-Motorized | PVT | 210 | 84 | 5 | 2 | 7 | 3 | 10 | 4 | 16 | 6 | 3.3 | 249 | 7 |
| Recreation | Total | 3099 | 89 | 48 | 1 | 265 | 8 | 35 | 1 | 22 | 1 | | 3470 | |
| TOTALS | | 14400 | 76 | 1446 | 8 | 2122 | 11 | 208 | 1 | 874 | 5 | | 19052 | |

Technical note: The table above was derived from a vector polygon coverage that did not have a minimum mapping unit applied. Hence, the GIS contains many very small polygons carried over from the raster classification. Thus, the average size of major change polygons as listed above is much smaller than if a minimum mapping unit of 1-2 acres had been applied to the analysis. Percentages are based on total acreage by ownership.

**TABLE 2. SAMPLE SECONDARY ANALYSIS:**
The full value of the type of data collected by this project becomes apparent when the above table is compared to goals and objectives presented in the Forest Plan. For example, the following table compares selected Forest Plan goals with the change detection results for three selected management units in the pilot area:

| MANAGEMENT PRESCRIPTION | SELECTED PLAN OBJECTIVES | | | CHANGE DETECTION RESULTS | | WERE GOALS MET? | |
|---|---|---|---|---|---|---|---|
| | Timber Man. System | Annual Creation of 0-9 Year Habitat | Retention of Open/ Semi-Open Areas | Major + Moderate Change per Year | Open/Semi-Open Retained | Creation of Young Stands | Retention of Openings |
| 3.1 Natural | Even-age | .7 to 1.5% of area | 4-10% of area | NF: 1.0%/yr | 1% | YES | NO |
| Vegetation | | per year | | PVT: 2.1%/yr | 16% | EXCEEDED | EXCEEDED |
| Communities | | | | ALL: 1.3%/yr | 6% | YES | YES |
| 3.4 Wildlife | Even-age | .8 to 1.5% of area | 7-12% of area | NF: 0.6%/yr | 2% | NO | NO |
| Habitat | | per year | | PVT: 1.1%/yr | 28% | YES | EXCEEDED |
| Emphasis | | | | ALL: 0.8%/yr | 11% | YES | YES |
| 6.1 Semi-Prim. | Uneven-age | 0 to 1.0% of area | 1-5% of area | NF: 0.1%/yr | 1% | YES | YES |
| Non-Motor. | | per year | | PVT: 1.5%/yr | 2% | EXCEEDED | YES |
| Recreation | | | | ALL: 0.2%/yr | 1% | YES | YES |

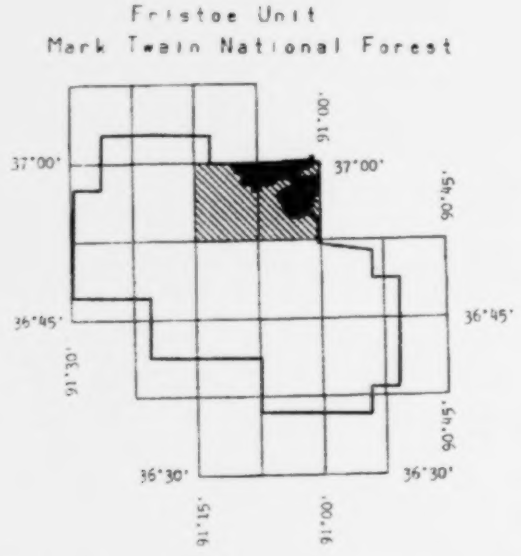

Fristoe Unit
Mark Twain National Forest
37°00'
37°00'
91°00'
90°45'
36°45'
36°45'
91°30'
90°45'
36°30'
36°30'
91°15'
91°00'

**Figure 2.** Map showing the boundaries of the project area and location of the boundaries of the three management units (shaded) used in the primary and secondary analysis.

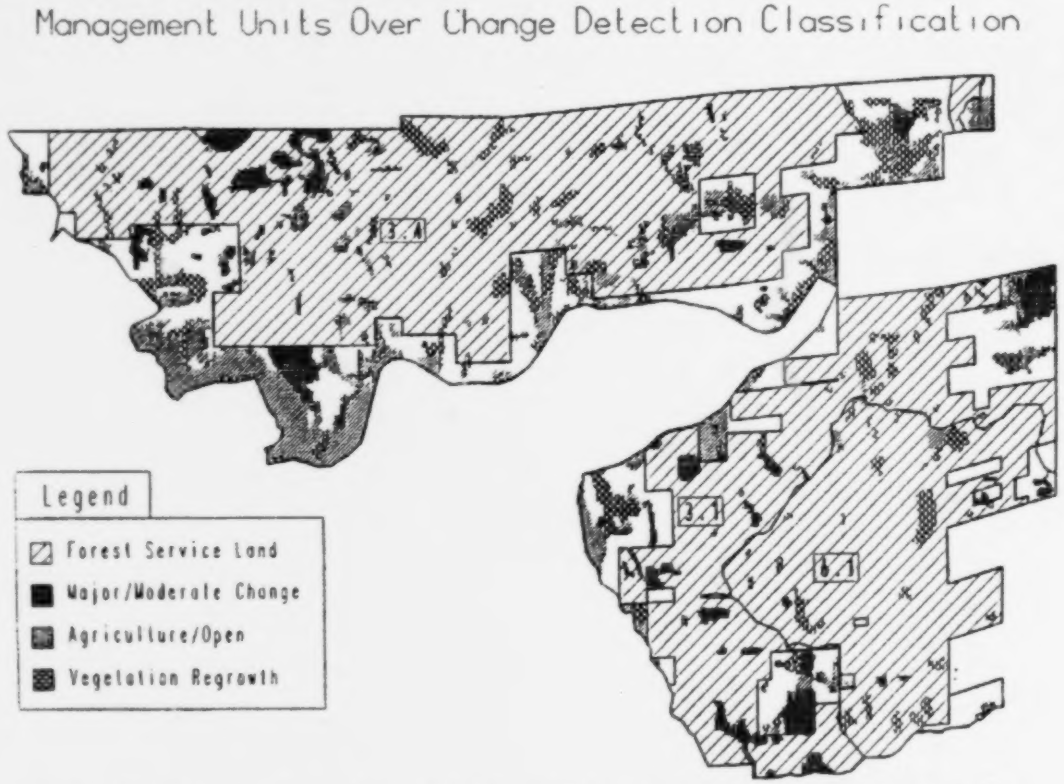

Management Units Over Change Detection Classification
3.A
3.1
6.1
Legend
▨ Forest Service Land
■ Major/Moderate Change
▩ Agriculture/Open
▦ Vegetation Regrowth

**Figure 3.** Map of the three management unit boundaries illustrating the results of the Landsat TM classification with overlay of land status.

layers of ownership and management unit prescription, serve as input to a number of other analysis and modelling processes. Tables 1 and 2 illustrate a these ideas.

## Analysis of the overlay process

Table 1 provides a sample analysis of three management units with different management prescriptions. The number of acres and percent of total land area for each type of activity were created by overlaying a GIS layer containing the management unit boundaries for three selected units over the change detection GIS. The additional layer of land status allowed for summary of activity on private versus public lands. Table 2 provides a secondary sample analysis that builds upon the results of Table 1. This table compares the goals and objectives presented in the Forest Plan to the results of the change detection summary. The full value of the data becomes apparent in this type of analysis.

Figures 2 and 3 show the extent of the project area and the three management unit boundaries used in the sample primary and secondary overlay analysis. Figure 3 illustrates the change detection map created from the digital Landsat TM classification for the three management unit boundaries. The numbers inside of the management unit boundaries refer to the management prescription in Table 1 and Table 2.

## Conclusions

Based upon the results of the Landsat TM classification and the GIS overlay analysis, the following conclusions were reached:

> Forest canopy changes can be accurately detected andclassified in a straight-forward procedure using multitemporal Landsat TM images.
>
> The derived TM classifications can greatly aid in the monitoring of forest related activities over time and the results can be readily integrated into a GIS.
>
> The results of the change detection become most effective when compared to the goals and objectives presented in the Forest Plan. Results can be tabular or maps.

## References

Colwell, J., G. Davis, and F. Thomson, 1980. Dectection and Measurement of Change in the Production and Quality of Renewable Resources. US Forest Service, Nationwide Forestry Application Program. AgRISTARS Report No. 145300-4-F.

Colwell, J., P. Sanders, G. Davis, and F. Thomson, 1981. Data Base Manipulation for Assessment of Multiresource Suitability and Land Change. US Forest Service, Nationwide Forestry Application Program. AgRISTARS Report No. RR-E2-04293.

Lee, Linda K., 1978. Interpreting Land Use Change Through Satellite Imagery. US Department of Agriculture Economics, Statistics, and Cooperatives Service. Report No. 442.

Malila, William A., 1980. Change Vector Analysis: An Approach for Detecting Changes with Landsat. Machine Processing of Remotely Sensed Data. ERIM Proceedings, Ann Arbor, MI.

Nelson, Ross, April 1982. Detecting Forest Canopy Change Using Landsat. Agristars Report No. RR-G2-04258.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Replacing aerial photo interpretation in resource inventories with integrated digital images and GIS

**Frank Hegyi**
(presenting author)

**Paul Pilon**

**Penny Walker**

Ferihill Technologies Ltd.
1025 Pandora Avenue
Victoria, B.C. V8V 3P6

## Abstract

*Information acquired for resource management through current operational methods lacks some of the critical data needed for the protection and enhancement of the environment. In particular, conventional aerial photographic techniqes are limited mainly to visual observations of the earth's surface. Even when satellite imagery is used, the resolution of the data does not accommodate the monitoring of individual trees and stands, especially during their early renewal development. Also, the existing technology has limited capabilities in demonstrating the impacts of resource management practices on the environment. For example, the use of video images for the evaluation of Visual Quality Objectives (VQO) may not be cost-effective when dealing with large areas.*

*Ferihill Technologies Ltd. is developing, in cooperation with Canada Centre for Remote Sensing, National Research Council of Canada and Forestry Canada, techniques to replace conventional aerial photography and video imaging in resource management with airborne digital data. Information collected with sensors capable of examining the earth's surface from a wider range of the electromagnetic spectrum than that of the visible bands, is integrated with 3-D thematic GIS information to provide unique cost-effective analysis techniques. This new approach is especially designed to collect and process information which is both sensitive and responsive to environmental issues in forest management. In addition, the system contains a wide range of tools to present the results of alternative management strategies including the generation of diagrams, graphs and images with their descriptive text.*

*In this paper, an overview of the new Airborne Digital Data Information System (ADDIS) is presented and its unique capabilities are illustrated through a slide presentation.*

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Managing forest resources with raster tools

**Andrew S. Bury**

ERDAS, Inc.
2801 Buford Highway NE, Suite 300
Atlanta, Georgia 30329 USA
Tel: 404/248-9000
Fax: 404/248-9400

## Abstract

*Image processing techniques, in conjunction with raster GIS modeling and vector data overlays, are being used to develop successful forest management practices throughout the United States. Ecological issues concerning timber resources, wildlife habitats, and watershed protection are seriously affected by forest management practices. Remote sensing and GIS offer an integrated approach to forest management for both publicly and privately held lands. This paper focuses on the process of managing timber sale compartments using satellite imagery and scanned aerial photography. Satellite data are compared to aerial photography for updating timber stand boundaries. This process was used to define impact-free stands needed to protect Red-Cockaded Woodpecker habitats and to improve tree stand updating and harvesting procedures.*

## Introduction

Image processing and geographic information system (GIS) modeling tools provide an integrated approach to forest management. Together, they can be used to support a variety of mapping and resource management activities. The highly accurate information generated through image processing and GIS techniques not only allows foresters to more efficiently manage their resources, but provides information that acts as a common ground between forestry companies and environmentalists. This information allows opposing groups to assess and resolve issues regarding the preservation of critical forest stands or wildlife habitats. Several recent projects are described in this paper to illustrate how image processing, raster GIS modeling, and vector data analysis can be combined for tree stand mapping and wildlife habitat protection.

## Updating forest stand maps using satellite imagery

One project, recently conducted by ERDAS, Inc., for a large forest management company, used satellite imagery to update out-of-date forest stand maps. In the past, the company had delineated forest stands on unrectified aerial photographs, then digitized them into their vector GIS. This technique generated a GIS data base that was not accurate, either in terms of the actual location of stands or in stand size. The inaccuracies caused errors in assessing the company's resources, especially for tax and harvest purposes. In addition, map updating was difficult because the updates took the form of hand-drawn maps done by foresters in the field. The GIS attribute data base, however, contained valuable information on tree stand characteristics. An important goal of the project was to preserve the attribute data from the company's original stand data base, while increasing the overall positional accuracy of the mapped information. The company also needed an efficient technique for updating old stands as well as adding new stands.

ERDAS acquired 10 meter SPOT panchromatic imagery, which was rectified and georeferenced. The tree stand maps previously digitized from unrectified aerial photographs were overlayed on top of the imagery, and warped to fit it. New stands that were readily apparent on the imagery were digitized directly from the computer screen. Others that were more difficult to detect on the imagery were annotated on image plots and digitized from a digitizing tablet. Changes in the boundaries of previously mapped stands were also rapidly updated using the same process.

Changes in the attributes of tree stands were made from field notes typed into on-site computers by foresters while conducting their forest management activities. These changes were uploaded to a central system where map updating took place. When a particular forest stand was updated, the system automatically searched for the corresponding attribute information and recorded the update in the stand data base.

Map updating with satellite imagery brought the company's inventory of forest resources up-to-date quickly and efficiently. It also presented a means for maintaining the data more effectively. Satellite imagery provided the most cost-effective and accurate base for updating maps of the company's entire land holdings at one time. Following the initial update, aerial photography was used to record changes in timber stands occurring in areas of high logging activity. Paper copies of the imagery and aerial photographs were provided to the field personnel on which to record changes as they occurred.

## Protecting Red-Cockaded Woodpecker habitats

A project recently conducted by graduate students in Landscape Architecture at Louisiana State University identified individual tree stands that should not be harvested because of their proximity to the habitat of the Red-Cockaded Woodpecker. The woodpecker, which is protected by law, favors a certain combination of forest species. This project focused on determining the impact the woodpecker habitats would have on USFS forestry management plan.

First, U.S. Forest Service (USFS) compartment and stand maps of a section of the Kisatchie National Forest near Alexandria, Louisiana, were digitized. These maps delineated homogeneous forest vegetation species. Because of the diversity of the area, the maps were highly complex, showing 382 distinct areas of forest vegetation in an area less than 25 square miles in size. USFS attribute data on tree stands, based on the U.S. Department of Agriculture Silvicultural Examination and Prescription Field Book, were imported into an ERDAS system. The attributes included forest type (e.g., Hemlock-Hardwood), and wildlife codes, among many others. The wildlife codes were used to create a file of stands that would typically include the Red-Cockaded Woodpecker.

Proximity searches of 0.25, 0.5, and 0.75 mile ranges were generated around the woodpecker habitats. The 0.25 mile buffer encompassed woodpecker foraging areas, whereas the 0.75 mile buffer represented its flight radius. The buffer zones were intersected with the USFS forest stand maps to identify any tree stands that at least partially crossed woodpecker habitats.

A software program was developed to read the histograms of the stand-buffer intersection files and create three new attribute fields. These fields indicated which stands intersected with which buffers. Acreage statistics were also generated, which indicated how many acres of each stand intersected with each buffer.

The new attribute fields were used to generate three images which showed the stands that intersected with each buffer zone. The maps and the accompanying acreage attributes provided USFS personnel with useful information to conduct forest harvesting studies, determine which stands and how many acres of timber could be harvested safely without threat to the woodpecker habitats.

## Mapping old growth stands in the Pacific Northwest

Pacific Meridian Resources completed a project for the USFS to produce GIS data on old growth forests in all national parks and national forests in western Washington and Oregon. Old growth forests are critical because of their species diversity and their role in providing a habitat for wildlife. Unfortunately, old growth is defined differently from person to person. In general, however, the definition centers on the following four characteristics:

Crown Closure - the percentage of ground area that is covered by the tree canopy.

Tree Size Class - measured according to the tree diameter at breast height.

Tree Species - identified when the canopy of a particular species occupies more than 25 percent of cover.

Stand Structure - either even storied (all trees approximately the same size and age), two-storied (two distinct generations), or multi-storied. Most old growth forests are multi-storied, supporting trees of many sizes and ages.

By mapping each of these characteristics separately, the USFS would have the flexibility to change the definition of old growth as needed and recalculate the extent and location of old growth forests. Additional information that was mapped and added to the USFS GIS data base included slope, aspect, elevation, hydrology, flight line maps, plant associations, location of research and inventory plots, and training site locations.

Pacific Meridian used a unique combination of GIS, image processing, and statistical analysis procedures on Landsat TM imagery to analyze variations in spectral reflectance and vegetation. Training sites were captured from the imagery and passed to the statistical analysis package to identify variables (e.g., aspect, elevation) which significantly affected vegetation distribution. Both supervised and unsupervised image classification were performed, and compared. The analyses resulted in four classification maps, one for each old growth vegetation characteristic (i.e., species, size class, stand structure and crown closure).

A set of spectrally and informationally unique training signatures were also generated, along with an identification of training signatures with confusing spectral reflectance. For example, spectral confusion typically occurred between similar species (e.g., Sitka Spruce and Pacific Silver Fir). These areas were further analyzed to determine if ancillary data (such as aspect, elevation, or latitude ) could be used to stratify the imagery. Fortunately, the distribution of species was usually found to be a function of these ancillary data.

The four old growth vegetation characteristic maps were combined to create a vegetation type polygon map. Polygons were delineated around areas of homogeneous vegetation using algorithms which replicated the decision rules used by photo interpreters delineating vegetation types. First, the crown closure pixel classification map was generalized to produce crown closure polygons. Crown closure was then held constant, and size class was generalized within each crown closure polygon. Next, both crown closure and size class were held constant and stand structure was generalized. Finally, species were summarized within each crown closure-size class-stand structure polygon.

The vegetation polygon map was overlaid with each of the four pixel classification maps to produce an attribute file. The attribute file described the distribution of species, size class, structure, and crown closure within each vegetation polygon, and provided an indication of the vegetation diversity of each polygon.

The old growth mapping projects required a classification accuracy of at least 80 percent. Pacific Meridian and Dr. Russ Congalton, a consultant from the University of California at Berkeley, developed procedures and a sample design for map accuracy assessment. Guided and random procedures were generated to compare manual photo interpretation with results obtained through the above procedures. At least 20 percent of the areas assessed for classification accuracy were ground-checked in the field. An error matrix was generated for each vegetation characteristic.

The use of the mapping procedures described above lowered the cost per acre of obtaining accurate information by one-third, and produced fourteen layers of detailed GIS data. Traditional air photo interpretation techniques would have yielded only one layer of data. The pixel classification maps, and vegetation polygon map with its attribute file provided the means by which users could query and model the data for forest management practices. These enabled the USFS to address issues such as the fragmentation of old growth and its implications to wildlife habitat, and to develop estimates of the biological diversity of forest vegetation.

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Using ancillary data in post-classification modelling to increase the accuracy of conifer species classification from Landsat Thematic Mapper (TM) data

**Michael Golden**

**Lisa Lackey**

Pacific Meridian Resources
Suite 570,
5200 S.W. MacAdam Avenue
Portland, Oregon 97201

## Abstract

*A post-classification model was used to identify and correct mis-classifications of forest tree species from Landsat TM data of US National Forests in Western Oregon and Washington. The model used elevation and aspect data, and Forest Service ecologists' expertise and plot data, to identify pixels with species labels outside their "allowable" elevation/aspect ranges. These pixels were relabeled with a species label from historical Forest Service maps of cover type. Separate models were developed for each National Forest. Manual editing after the model was run was used to further improve the map. The Mt. Hood National Forest model is presented as an example.*

## Introduction

Pacific Meridian Resources, under contract to Region 6, USDA Forest Service, produced a Geographic Information System (GIS) data base for old growth on national forests in western Oregon and Washington (Teply and Green, 1991). The old growth project was part of a larger project to produce a GIS data base for all the National Forest lands in Washington and Oregon. This data base includes 15 layers of information and represents more than 30 million acres. The old growth study involved 12.5 million acres.

Vegetation layers of tree crown closure, tree size class, forest stand structure, and tree species were produced using Landsat Thematic Mapper (TM) satellite data. This paper presents an example of using ancillary data in a GIS modeling process to improve the computer classification of TM data for tree species. The Mount Hood National Forest is used as an example. Similar models were developed for each of the nine national forests in the old growth mapping project.

## Background

The success of image classification is based on the assumption that there is a correlation between the thematic classes of interest and the spectral response of those classes. Unfortunately, this correlation is not strong with respect to conifer tree species. While some tree species in the Pacific Northwest (most notably lodgepole pine, *Pinus contorta*) have a distinct spectral response, most species and species mixes have very similar spectral responses. In order to accurately classify tree species, additional data was needed to differentiate species and species mixes to the level desired for this project. Fortunately, species distribution in the Pacific Northwest is highly correlated with elevation and aspect. We used this relationship to improve the classification of the imagery.

Ancillary data can be defined as "secondary data pertaining to the area or class of interest, such as topographic, demographic, or climatological data" (Swain and Davis, 1978). In a remote sensing context, ancillary data refers to any data not obtained by the sensor. Ancillary data, such as topographic data, can be used either before, during, or after classification, through stratification, classifier operations, or post-classification sorting and post-classification modeling (Hutchinson, 1982).

Stratification is a technique to divide a large data set into smaller homogeneous subsets. Separate classifications can then be run on each subset. The objective is to reduce the variability within each data set. However, stratification divides the data into subsets with distinct boundaries. No gradations, such as those found in nature, exist between mapping classes (Hutchinson 1982), and "seam lines" are often apparent when the subsets are "stitched" back together.

A second use of topographic data is during classification. Digital terrain data can be used as an additional "band" of data, called a "logical channel" (Strahler et al. 1978). This use of terrain data has been disappointing. The "simple addition" of non-spectral data appears to add little to classification accuracy (Hutchinson 1982). Topographic data can be used to modify prior probability in the maximum likelihood classifier (Strahler 1980). If the distributions of cover types across the topography are known, the probability of occurrence can be used to "weight" the maximum likelihood classifier. Strahler et al (1978) reported that classification accuracy was improved 13 to 19 percent using topographic data to modify the maximum likelihood classifier. Hutchinson (1982) discusses two problems with using modified probabilities. The assumption of a normal distribution is often not met for ancillary data. Secondly, determining prior probabilities requires more intensive ground sampling and therefore increases costs.

A third major use of topographic data in image classification is for post-classification sorting, usually with unsupervised classification. In a study in California, Woodcock et al. (1982) found that "due to varying slope angle and aspect, poorly-stocked open stands of conifers under low angles of illumination exhibited spectral signatures which were very similar to those of dense, well-stocked stands under more direct illumination". They used digital terrain data to produce a map of illumination. Overlaying this map on the clusters divided all the classes that were spectrally similar but differed in illumination. This allowed the analysts to separate poorly-stocked stands from well-stocked stands. Post-classification sorting suffers from its deterministic nature (Hutchinson 1982). It uses data that is essentially continuous, but produces classes based on abrupt criteria boundaries.

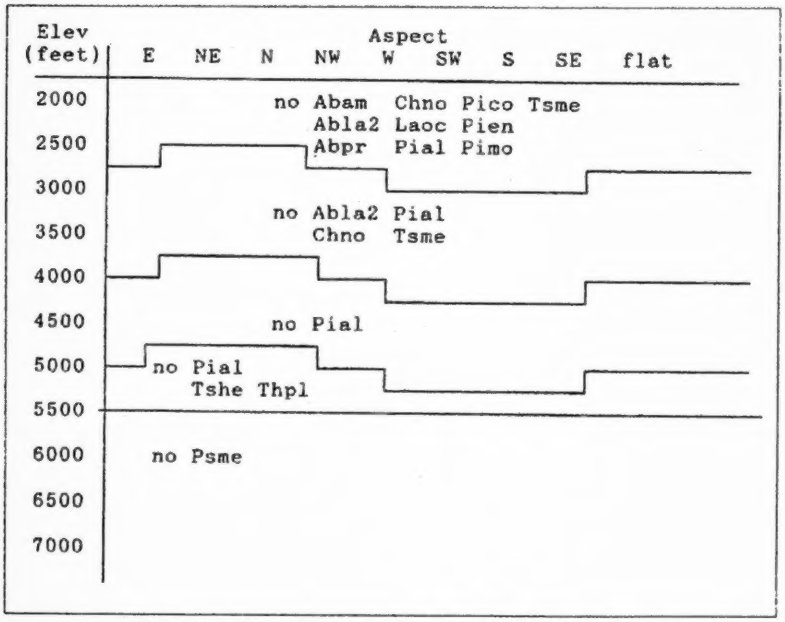


Figure 1. Aspect/Elevation ranges for Mt. Hood tree species.

In this study we used a fourth method of incorporating ancillary data, post-classification modeling. Topographic data and knowledge about their relationship to tree species distributions were used to identify problem areas in the classification. Historical cover type maps and manual editing were used to correct these areas. This approach avoided the "seam lines" inherent with pre-classification sorting, and we felt we did not have adequate information for estimating prior probabilities. This was not a research project and so there was no hypothesis to test. Rather, it was an attempt to incorporate available information into the classification to improve the accuracy of the map.

## Data used

The primary data source for species classification was Landsat TM imagery, geocoded and terrain corrected to UTM coordinates. Digital Elevation Model (DEM) data, ortho photo quads, and resource aerial photography were also available. Field data included representative training sites established by Pacific Meridian Resources staff and Forest Service ecology plot and inventory data.
All training site and plot data included information on species composition, aspect, and elevation. Pacific Meridian Resources staff also recorded extensive field notes on ortho photo quads while driving between training site locations. Forest Service historical maps dating from the early 1950's were also available for each national forest. These maps were digitized and converted to raster format using ESRI ArcInfo (Redlands, California) software.

## Methods

A combined supervised/unsupervised approach was used to classify the TM imagery. All classifications were performed using ERDAS (Atlanta, Georgia) image processing software. For the old growth study, only those forested areas meeting the tree density, size, and stand structure criteria for old growth were mapped for species. Forested areas not meeting those criteria were mapped as "young conifer". A particular species was included in the species label if it represented at least 25 percent of the crown closure of the stand, up to 3 species.

Based on Forest Service plot data and training site data, aspect/elevation ranges for each major tree species were developed. After consulting with Forest Service ecologists, these ranges were combined and generalized into one aspect/elevation matrix which delineated ranges where certain species were *not* allowed (see Figure 1).

An ERDAS gis model was written that compared, pixel by pixel, the species label in the original classification with the aspect and elevation and the allowable range for that species. In this way, "unreasonable" species calls (i.e., species calls outside their allowable ranges) were identified. The model did not insure accurate species labels, but rather identified labels which were not allowed. "Unreasonable calls" then had to be replaced with an acceptable species label. Both Forest Service ecologists and Pacific Meridian Resources staff agreed that the historical species map would be a reasonable substitute for the mis-classified labels. It was assumed that species composition in the old growth stands would not have changed appreciably over time. However, the historical species labels were less detailed than the classification label and so were not desired for the entire coverage.

The final step in producing the species coverage was to manually edit the coverage based on plot data, training site data, and field notes from the ortho quads. The editing was completed for all the forests on a quad by quad basis.

## Results

Results of classifying, modeling, and editing species on the Mt. Hood National Forest are summarized in Table 1. The Mt. Hood model identified 77841 pixels (1.5%) labelled with species outside their allowable ranges. These pixels were relabelled with the historical species call. Manual editing relabelled an additional 790910 pixels (14.8%). These changes are assumed to be a more accurate species label. Similar results were obtained with the models for the other National Forests.

|  | % of pixels classed by original classification | % of pixels classed by model | % of pixels classed by editing |
|---|---|---|---|
| Original Classification | 100.0 % |  |  |
| After Modeling | 98.5 % | 1.5 % |  |
| After Editing | 83.7 % | 1.5 % | 14.8 % |

Table 1. Modeling and Editing Summary.

Quantitative accuracy assessment is currently being implemented on each of the individual vegetation layers in the old growth data base. At this writing, no results are available for the species layer.

## Conclusion

A post-classification model was used to improve the image classification of tree species on national forests in Western Oregon and Washington. The model used DEM data, field data, and personnel expertise on species distribution. The model identified species labels which were "unreasonable" with respect to aspect and elevation. Unreasonable calls were replaced with a label derived from historical maps of species distribution. After modeling, the species maps were manually edited.

The Mt. Hood model relabeled only 1.5% of the pixels. While this is a small percentage of the total number of pixels, we feel that this is an important improvement. Geographically, they were found in both areas where we had field notes and data, and in areas where we did not. Those pixels that changed were clearly misclassified. From a remote sensing perspective, those errors are easily understood. However, correcting the pixel labels which should not occur is an important step in insuring acceptance of a map by clients with little or no remote sensing background. The absolute percentage of corrections may be less important than the type of corrections made.

Manual editing relabeled 14.8% of the pixels. Manual editing is a critical step in commercial mapping applications of satellite imagery. We must understand, and be willing to admit, the limitations of imagery for mapping purposes. We can then use ancillary data to improve the map. Manual editing allows us to use site specific information that would be difficult, if not impossible, to model.

The Mt. Hood model was based on the data available and was relatively simple to implement. While it did not guarantee correct species labeling, it eliminated labels which were clearly incorrect. Manual editing further improved the accuracy of the map. Obviously, the model is limited by the data used to derive it. The model did not consider small micro-climate influences such as cold air pockets, or precipitation/moisture gradients. However, the model was a useful tool in improving the classification of forest species.

## References

Hutchinson, Charles F. 1982. Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement. Photogrammetric Engineering and Remote Sensing, Vol. 48, No. 1, pp.123-130.

Strahler, Alan H. 1980. The Use of Prior Probabilities in Maximum Likelihood Classification of Remotely Sensed Data. Remote Sensing of Environment 10:135-163.

Strahler, A.H., T.L. Logan, and N. Bryant. 1978. Improving Forest Cover Classification Accuracy from Landsat by Incorporating Topographic Information. *In: Proceedings of the 12th International Symposium on Remote Sensing of the Environment, Vol. II,* Environmental Research Institute of Michigan, Ann Arbor, MI, pp.927-942.

Swain, P.H. and S.M. Davis (Editors). 1978. Remote Sensing: The Quantitative Approach. McGraw-Hill, Inc., 396p.

Teply, John and Kass Green. 1991. Old Growth Forests: How Much Remains? GeoInfo Systems. Vol.1,No.4. Aster Publishing Corporation, Eugene, Oregon. pp.22-31.

Woodcock, C.E., J. Franklin, and A.H. Strahler. 1982. Improvements in Forest Classification and Inventory Using Remotely Sensed Data. *In: Proceedings of the 16th International Symposium on Remote Sensing of Environment, Vol. II,* Environmental Research Institute of Michigan, Ann Arbor, MI, pp.963-974.

# Human Resources Working Group

**Hardware, software and flesh: a human resources architecture for geographic information systems**

Robert Menes, Ministry of Environment, Lands and Parks, Victoria, BC, Canada

**New GIS/LIS education and training initiatives in Australia and southeastern Asia**

Stephen R. Kessell, School of Computing, Curtain University of Technology, Perth, Western Australia

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Hardware, software and flesh: A human resources architecture for geographic information systems

**Robert Menes, P.Ag.**

Senior Land Information Analyst
Surveys and Resource Mapping Branch
Ministry of Environment, Lands and Parks
4th Floor, 1802 Douglas St.
Victoria, B.C. V8V 1X4, Canada
Tel: 604-387-9353
Fax: 604-356-7831

## Abstract

*The Land Information Infrastructure (LII) of British Columbia is being designed to produce a system for data access and interchange by the year 1998. The system architecture, precursor to the high level design, identified the logical components of the system from a hardware and software point of view. However, the architecture and design must also include the definition and management of the human resources necessary for system operation. Identification and management of the human resources is complicated by the fact that the organizational structures for managing people do not map onto the system structures for the Land Information Infrastructure. To do so, an additional architecture must be defined which deals with the human resources necessary to assume responsibility for the computer/communications system. This paper presents a human resource architecture which can be used for the operations and management of the LII and presents a possible implementation of the architecture. The architecture permits the development of plans for realizing the operations team and facilitates the estimation of resource needs for the system.*

*In the paper, responsibility categories, assigned to people, are defined which are necessary for the operation of the system, such as: spatial data management, internal user management, policy and human resource coordination. Every responsibility category inherits the operations of a management class: planning, resource allocation, reporting, and monitoring. The definition of the classes, and their interrelationships, form a model for human resource responsibilities. Once the model is defined, it is applied to British Columbia's Land Information Infrastructure to illustrate the methodology for identifying the number of people needed to manage a system and their responsibilities.*

## Introduction

### Motivation

Geographic Information Systems are becoming essential tools in the management of land-related information throughout North America. People responsible for making land-use decisions at all levels are aware of the new technology and are eager to make it part of the standard set of management and analysis tools. However, the implementation of GIS is far from simple and, regardless of the scale of the implementation, key issues arise which cannot be solved by the application of technology alone. GIS requires the management of multiple sets of data, spatial and nonspatial, linked together in various ways. The underlying software and hardware platforms must be matched and the communication with the "outside" world must be defined. In short, GIS places serious management requirements on the use of the support tools.

GIS is not the only technology which impacts the modern planner and policy maker. Communications networks enabling distributed processing and distributed data storage are making disparate data sources accessible. Technologically, they are accessible, but managing the information across regional and administrative boundaries, across systems, is an unresolved nightmare.

Historically, systems design has been performed for the "technical" components of a system: the hardware and software and communications networks. After the system is designed, the human aspects of the system are considered: how it will be run, maintained and redeveloped. To date, there is no clear methodology for identifying the human resources necessary. Usually, standard job titles such as "data administrator" or "systems operator" are used post hoc to assign the human resources. The management of the human resources and the planning for system growth is often not defined.

British Columbia is constructing a system for sharing land-related data within all of the ministries in the government. The system spans ministries, geographic regions, data types, hardware and software platforms, and user levels. The extensive system will be created over a seven year period in four phases. At this time, Phase I is nearly complete. A set of requirements has been identified, a target system and migration strategy developed, and a logical model of the working environment has been created. The system, known as the Land Information Infrastructure, is already defined by a System Architecture and a High Level Design. Phase II entails the design and creation of a prototype system. A fundamental piece of the design is missing: the human resources necessary to manage the system. Is the Government of British Columbia creating a system which cannot be operated and maintained? A means of identifying and assessing the human component of the system is needed. This paper attempts to address this requirement.

## Objective

The objective of this paper is to provide a means of defining the human resource needs of the British Columbia Land Information Infrastructure (LII). The LII is more than a Geographic Information System: it is a large scale network of services, provided by hardware, software, and data, to facilitate the sharing and exchange of data across the province. A system architecture has been defined, but the full scale system has not been designed in detail. This paper adds a human resource architecture to the systems architecture to provide a basis for the design of a system which includes, explicitly, the people needed to run it.

The objective of the architecture is to define the human resources for the B.C. Land Information Infrastructure such that increases in the volume or size of the system require only a scaling of the human resources allocated to manage the system.

A secondary objective of the paper is to provide a model for human resource systems design for a) general systems development and b) geographic information systems development.

## Methodology

The methodology used to define the human resources architecture relies on identifying a model for the responsibilities of the people required to manage a computer-based data exchange system. Common terms for job responsibilities will not be used to avoid confusion with conflicting definitions.

An architecture is a plan for the design of a system which defines the types of system components and the relationships between the components. It is general in nature and usually relates to logical entities rather than physical objects. For example, a human resource architecture could include an entity that manages all data in the system without identifying whether that entity is one person or a group of people. Use of the architecture involves defining how these logical entities map onto real physical objects.

The human resource objects — those entities which perform the functions usually assigned to people — are identified first. While broadly defined, these objects provide a way of making distinctions between types of responsibilities to be assigned. One person may be assigned more than one of these human resource objects. The objects are then further refined, and additional objects will be created which are subclasses of the more general objects. Objects can inherit behaviour from their parent class. The result of the human resource object definition is a class hierarchy of objects which spans all of the human activities required to manage the system.

The human resource objects must be defined in terms of their functions (responsibilities), protocols (communication with other objects), and structure (parts or attributes). Once this has been done, we have a complete set of entities which can be used to form a model. A model will be formed which identifies how the human resource objects relate to each other and the world external to the system.

The last phase of the methodology is the application of the model to a specific system architecture. This entails mapping the components of the system architecture onto the human resource objects in this model. One additional step is required to implement the model: the human resource objects must be assigned physical values (levels of effort in man-years) to define the actual human resources needed. The process is depicted in Fig. 1.

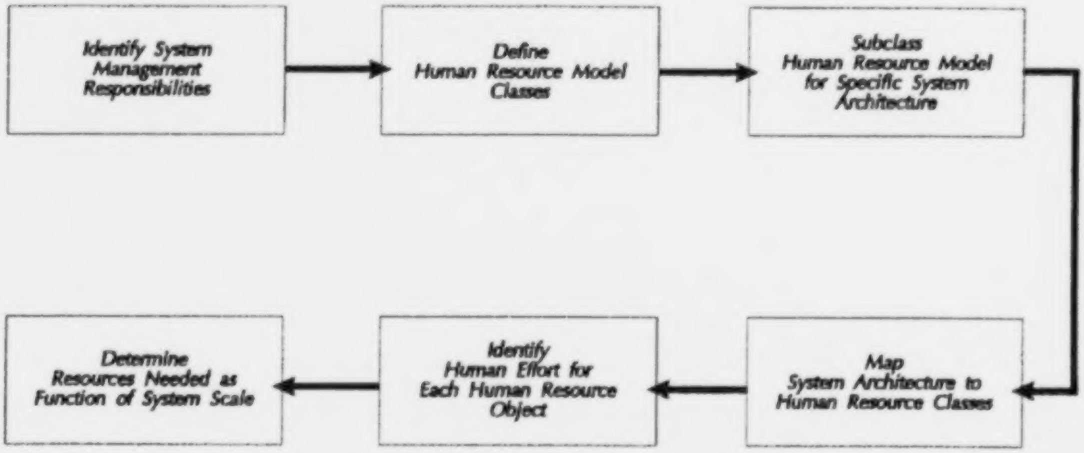

Identify System Management Responsibilities
Define Human Resource Model Classes
Subclass Human Resource Model for Specific System Architecture
Map System Architecture to Human Resource Classes
Identify Human Effort for Each Human Resource Object
Determine Resources Needed as Function of System Scale

*Fig. 1. Human resource management methodology*

## The human resource management model

### The System

The system which must be managed is composed of hardware, software, and information (data). The hardware runs the full range from personal computers to mainframes. The software includes both operating systems and data-specific applications. The data includes internal system information (accounting, usage, etc.) as well as the spatial and nonspatial information which is the main commodity of the system.

### Model requirements

The model used to describe the human resources of a geographic information system must meet the following requirements:

- simple: It can be understood by a wide range of planners and systems architects
- general: It can be applied to systems with complex components of varying types
- decomposable: It can be broken down into its parts and analyzed in discrete pieces
- specifiable: It can be described unambiguously and built into a computer-based system
- implementable: It can be applied in a practical situation
- organized: It can be applied and described in linear, logical steps

Furthermore, it must be able to address the following key issues:

- Common management functions must be able to be described and recognized by system planners
- Specific lines of communication must be able to be described to eliminate confusion in the assignment of responsibilities
- Communication with both man and machine must be specified to acknowledge the fact that management must be performed for both the human and computer systems
- Implementations will vary in complexity and the model must be able to respond accordingly
- Widely disparate responsibilities may be combined in an implementation. One person may assume responsibility for many aspects of the system management
- External system interfaces must be identified to recognize the impact of the world outside of the system

### Basic model components

There are five basic components to the model which, when broken down further, are associated with human resource functions. These five components are:

- DataManagement
- TechnologyManagement
- ApplicationsManagement
- UserManagement
- MetaManagement

The first three management components relate directly to the three basic aspects of any computer-based system: data, technology and applications. The last two components add to the system the management of the users and the management of the resources for management, i.e. human resources themselves and the policy implemented by the system.

Every model component serves a management function within a particular domain. As such, there are functions - tasks - which must be done by each of these model components. These functions are the responsibilities of any management construct and include a) planning, b) resource allocation, c) reporting, and d) monitoring. There can be management of other people or of the hardware/software resources.

### *Data management*

Data is the primary currency of the system and requires rigorous management to ensure a common understanding, efficient access, optimal physical organization and storage, and integrity. Data Management for spatial data systems is particularly burdened by the need to manage the link between baseline geographic data and a host of nonspatial, or attribute data.

The Data Management aspect of the architecture encompasses all of the traditional database management functions as well as the data management functions [Martin, 1983]. Roles which perform these functions have been defined as Data Strategist (planning and co-ordination), Data Administrator (data analysis, definition of views and schemas, data modelling, validity, consistency and accuracy of data), Database Designer (physically structuring data and planning access methods, database backup and recovery, physical schemas, database performance, security, searching techniques, concurrency), and Data Operations Supervisor (data errors, system restarts and recovery, reorganization of databases and indexing, audit trails, computer scheduling, file transfer for export). In the ANSI/NDL model for database systems [Gorman, 1991], these roles fulfill themanagement of the dictionary and logical database, physical database and system control, and interrogation and system control (respectively). The physical database and system control aspects of database management are only partially performed under the DataManagement component. These aspects are part of the functions of the ApplicationsManagement and Technology-Management components.

DataManagement requires human resources which deal with the semantics, access, quality, and physical storage of the data.

### *Applications management*

Applications are the hardware/software systems which provide the specific services to the user. They are the business functions integrated into automated systems. An example of an application for the Land Information Infrastructure is the Repository, which holds and manages all of the metadata (data dictionary and directory information) for the system. Applications can be of two types, Data General and Data Specific. Data General applications, like the Repository, are not dependent upon the domain of the data which they utilize. Data Specific applications are designed to perform a set of tasks on data from a particular domain, e.g. forestry timber supply. For the LII, most applications identified in its High Level Design are Data General BC, 1991].

Applications Management involves the operation, maintenance, and redevelopment of the computer-based systems. The responsibility for the physical database and system control functions of database management sy-stems would fall under this component. In addition, the management of all of the input/output services of the system would be the responsibility of Applications-Management.

Clearly, there is a close interaction between Data-Management and ApplicationsManagement. However, ApplicationsManagement is more concerned with the "how" while DataManagement is focused on the "what".

### *Technology management*

Applications are built on top of technology platforms. These are the specific products and standards which can be used to create the applications which, ultimately, satisfy the user needs. Technology includes the computer systems, operating systems, input/output devices, communications networks, database management systems, geographic information systems etc. which are generic in nature and which must be combined with context-specific data to meet the needs of the system. Management of the technology involves the application of agency-wide policies to the purchase and implementation of the products, optimization and review of their performance, maintenance and upgrading, and migration to newer, more reliable systems.

### *User management*

Users interact with the system in terms of the data, applications, and technology. However, the interface to users and the management of their concerns is of such a fundamental nature that this set of functions should be extracted from those of the other components. User management entails the communication of user needs to the system management as well as the communication of system management changes to users. System training, information, and marketing fall under user management.

There are two types of users who must be managed as part of the system: internal and external. External users are users who request the data management and applications services of the system. This category can also be identified with "end-users". Internal users are the system operators and maintainers themselves: the human resources of the system management. These users require a different set of services and yet still require all of the management services provided for External users.

| Data Management | Applications Management | Technology Management | User Management | MetaManagement |
|---|---|---|---|---|
| Semantics | Operations | HW/SW Purchase | Training | Policy |
| Access | Maintenance | Implementation | Information | Human Resource Coordination |
| Quality | Redevelopment | Optimization | Marketing | |
| Physical Storage | | Migration | Feedback | |
| | | | Accounts | |

***Meta management***

Meta Management is the management of the human resources who manage and operate the system. There are two aspects to the metamanagement: human resource coordination, and policy. The first aspect is the traditional management function of personnel assignment, allocation, task definition, etc. The second aspect is the identification of procedures, standards, policies, etc., which must be implemented by the management of the system.

The five components are summarized in Table 1 above.

In addition to the five basic model components, there are also external components which require some interface. These components are also functional groupings identified with human resources:

- Users
- ApplicationsDevelopers
- DataProviders
- Vendors
- AgencyManagement

Users are the "end-users" of the system (discussed previously in the context of External users). ApplicationsDevelopers are those people or agencies who are contracted to provide specific applications for the system. DataProviders are those groups who contribute data to the system and, perhaps, manage that data independent of the system itself. Vendors are businesses providing commercial hardware and software for the technology platforms. AgencyManagement is the larger human resource planning authority to which the Human Resource Coordination must report.

Given the model components described in this section, we can present a simple model depicting the interaction of the these management entities.

The diagram shows the interactions between the main model components and external components as well as the interactions between each of the main components. These interactions will be described in greater detail in the following section. At this point, a complete set of objects which represent human resource responsibilities can be specified.

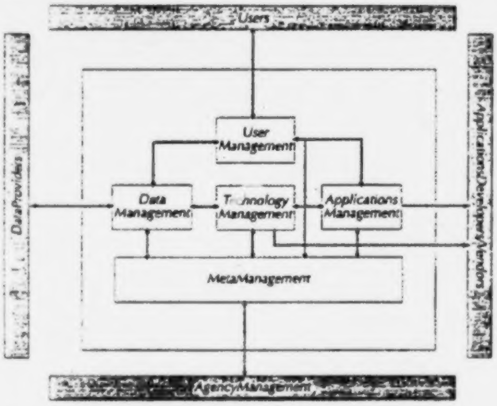


*Fig. 2. Human resource management model components*

## Human resource object hierarchy

The entities that represent the human resource responsibilities are complex: they share behaviour, they have compound structure, and they have subgroups which involve an additional level of specific behaviour. For these reasons, a hierarchy of objects which are divided into classes is an appropriate way of describing the components of the model. Once all of the classes of objects have been described, applying the model to a real system only requires that we instantiate the classes as needed.

The object hierarchy could be devised around the responsibility groupings or around the human management groups (e.g. people, or teams). The approach here utilizes the responsibility groupings. When the model is implemented, these responsibility groupings — the objects — are assigned to people. A number of objects could be assigned to a single person.

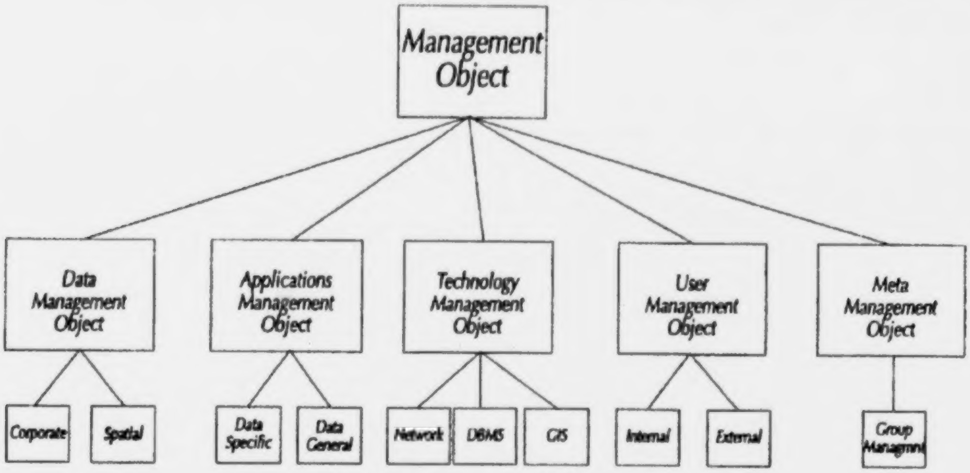


***Fig. 3.*** *Management model classes*

The human resource objects are grouped into classes. Each class inherits behaviour from its parent class. The behaviour of a class is its responsibilities, attributes, and protocols (the messages it responds to). For the purpose of this model, a class can be described by its parent class, a list of its responsibilities (functional areas), the other objects it responds to, and whether its communication is human-to-human, or human-to-machine, or both. The details of how the class responds to other objects is not presented here.

The table in the previous section listed the basic responsibilities of the major classes of the model. These responsibilities could be envisioned as attributes, or components, of every object of that class. As attributes, they may be composed of other objects with complex structures in themselves. This paper does not describe the complexities of the attributes of the major objects. However, a full implementation of the model would require a detailed description of all of the attributes of the classes.

The diagram above depicts the class hierarchy for the human resource management model. The root of all of the classes is the ManagementObject. The classes for Users, DataProviders, ApplicationsDevelopers, Vendors, and AgencyManagement are not shown and are considered to be external to the system.

## Description of object classes

The description of the five main classes was presented in the previous section. The parent class and their subclasses will be described herein.

### Management object

The root class of the hierarchy has the responsibility for planning, resource allocation, reporting, and monitoring. These tasks are complex in themselves and may be implemented by separate individuals. However, there should not be any instances (objects) of this class; it is simply a convenient way to extract the behaviour which is common to all management areas. All Management-Objects (and, therefore, all of its subclasses) respond to Meta Management Objects. That is, every management object must take direction from or provide information to the person who assumes the responsibilities for human resource coordination. In fact, that group of responsibilities is itself a subclass of Management Object and must respond to some Meta Management Object. The model is recursive. Note that Management Objects always communicate with other objects via human:human communications.

Subclasses of the Management Object have additional responsibilities which differentiate their behaviour. For example, the Data Management subclass adds the responsibilities for semantics, access, quality, and phy-sical storage of data to the planning, resource allocation, reporting and monitoring. One could view the responsibilities inherited from the Management Object class as methods, or functions, which must be applied to each aspect of the Data Management. The Data Management subclass must also respond to objects from other classes: ApplicationsManagement, Technology-Management, User Management, and Data Providers. Communication may take the form of human:human interaction (inherited from the Management Objectclass) or human:machine interaction.

The other four major classes similarly involve more specialized responsibilities and communication. If the model were to stop at this level of detail, however, it would be of little use: further subclasses are required to map this model onto a technical systems architecture. There are many ways to subclass the major model components and mapping the model onto an architecture involves the appropriate choice of subclasses. A number of subclasses can be defined which apply to large, spatial data systems:

### Corporate data management

This subclass of DataManagement adds the responsibility for all data which is directly managed by the central data infrastructure. Metadata, such as data definitions and schemas, would be corporate data in a distributed network system.

### Spatial data management

Geographic Information Systems which involve both spatial and nonspatial data require the implementation of specific techniques for managing the spatial data. The use of spatial data will span all applications in a distributed system and this subclass identifies the responsibility for managing this type of data.

### Data specific applications management and data general applications management

There are two types of applications which recur in a distributed spatial data system: those that operate on specific domains of data and those that operate on all data in the system, or system internal data. If there is a standard user interface to the database management system, this would be considered a data-general application. However, an application which calculates the area of a watershed, for example, would be a data-specific application. Each of these cases may be represented by a subclass of the Applications Management class.

### Network, DBMS, and GIS management

The TechnologyManagement class can be subclassed to divide the responsibility for the different aspects of a system. Once systems are designed along the OSI (Open Systems Interconnection) Model standards, subclasses could be defined for each layer of the seven-layer model [Judge, 1988]. However, at this time, the network, DBMS, and GIS aspects of the technology - which includes both hardware and software to support each function, may be an appropriate subdivision.

### Internal and external user management

The two types of users of the system discussed previously are convenient boundaries for the UserManagement subclasses. The actual ways in which communication will take place between the group or person responsible for UserManagement and these two types of users will differ greatly. However, the responsibilities of User Manage-ment would apply to both subclasses. Note that all User Management involves human:human communication.

## Application of the human resource management model to BC's LII

### Brief description of the LII

The Land Information Infrastructure of British Columbia has a system architecture [BC, 1991] which is designed to accommodate spatial and nonspatial data exchange and sharing across agencies and across the Province of British Columbia. An architecture has been designed which utilizes a) a distributed database, b) many levels of federated and local data schemas, c) client-server pro-cessing, d) integration of disparatetechnologies. The system architecture is depicted in the diagrams on the following page, in Figures 4 and 5.

The LII architecture includes a variety of logical components to facilitate both data and technology-integration:

- Network Query Server
- Repository Server
- Access Control Server
- Accounting Server
- Data Access Servers
- User Access Terminals (UAT's)

These servers and terminals provide access to the LII Repository (a shared directory and data dictionary), member databases, and applications of the LII (e.g. data viewers and translators). The data within the Repository consists of schemas which can be federated (spanning many databases with identical meaning) or external (single member database schemas). The UAT is an application which provides a single logical window into the system, both for users and external applications.

### Mapping the LII onto the human resource management model

An in-depth description of the LII System Architecture is beyond the scope of this paper. The essence of the problem is that the architecture contains a number of components which involve data, applications and technology, are distributed across the province, and which require some human operation. Mapping the system architecture onto the human resource architecture requires the identification of the components of the system architecture with objects represented by classes in the human resource architecture.

The mapping can be done in the following way:

a. Data Management objects are mapped to the responsibility for the data within the Repository, representing all of the schemas of the system. This data is uniformly corporate. However, each federation identified in the system design should involve the creation of a subclass of Data Management. Thus, there should be a Spatial Data Management subclass.

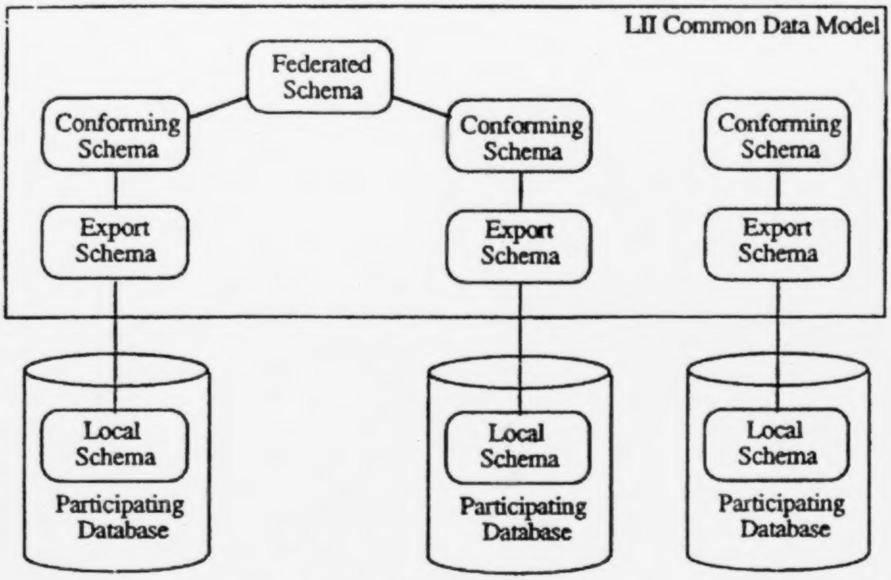

LII Common Data Model
Federated Schema
Conforming Schema
Conforming Schema
Conforming Schema
Export Schema
Export Schema
Export Schema
Local Schema
Participating Database
Local Schema
Participating Database
Local Schema
Participating Database

**Fig 4.** The LII data architecture

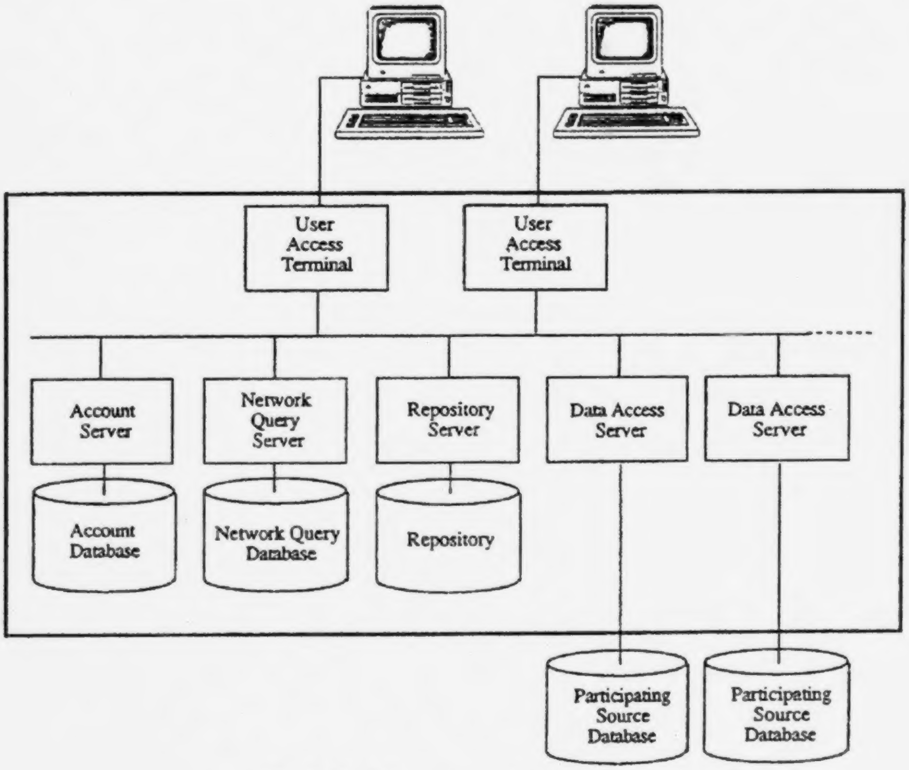

User Access Terminal
User Access Terminal
Account Server
Network Query Server
Repository Server
Data Access Server
Data Access Server
Account Database
Network Query Database
Repository
Participating Source Database
Participating Source Database

**Fig 5.** The LII component architecture

b. The LII does not have any data-specific applications at this time. Thus, there is no need for such a subclass. However, each server identified in the architecture represents an application which must be managed by one or more Applications-Management objects. Applications Management should be subclassed along the lines of the different applications, including: Repository Applications Manage-ment, Data Access Applications Management, UAT Management, and Internal Applications-Management. The last subclass aggregates the Network Query Server, Account Server, and Access Control Server applications. Each of these three subclasses have radically different environments to manage. In addition, expansion of the system can occur by the increase in the number of databases linked to the system, thereby increasing the human resources necessary for the Data Access-Applications Management, or by an increase in the physical infrastructure of the system (the computing power), thereby increasing the human resources necessary for the Internal Applications Management.

c. The User Management subclasses of Internal User Management and External User Management are applicable to the LII. The human resources necessary to assume these responsibilities will be a function of the number of internal and external users. In a similar way, the Meta Management responsibilities of human resource coordination will be a function of the number of instantiated Management Objects required by the size of the system.

d. Technology Management for the LII depends both on the physical demands on the system and on the changing external environment. The people responsible for the Technology Management must maintain the technical link with the world external to the LII and, as new technology becomes available, they must migrate the LII to take advantage of the changes.

The mapping of the system architecture onto the human resource architecture reveals that the human resources required to manage the LII can be determined primarily as a function of the external demands placed on the system. An increase in the number of users and data providers creates a direct increase in the human resources required for data, applications, and user management, and a secondary increase in the human resources required for the technology and meta-management.

**Further identification of human resources**

The bottom line for the allocation of human resources is the identification of the number of people needed to manage the system, and the specification of the skills required by those people. The human resources architecture presented enables the definition of both the number and types of people required. The LII requires the instantiation of objects of the following classes, listed below with their functional dependencies. (See Object Dependencies).

The only way to begin to identify the human resources required is to estimate the number of DataProviders and Users of the system and the human effort required to implement a single human resource object of a particular type. If we assume that each object corresponds to the effort of one person, then we can map this onto the load of the system. *For illustrative purposes only, the mapping on page ten will be used.* (See Object Equivalence).

| OBJECT | DEPENDENCIES |
|---|---|
| DataManagement | DataProviders |
| SpatialDataManagement | Users, DataProviders |
| CorporateDataManagement | Users, DataProviders |
| RepositoryApplications | Users |
| DataAccessApplications | DataProviders |
| UATManagement | Users |
| InternalApplications | Users |
| InternalUserManagement | Users |
| ExternalUserManagement | Users |
| NetworkTechnologyManagement | Users, Vendors |
| DBMSTechnologyManagement | Users, DataProviders, Vendors |
| GISTechnologyManagement | Users, DataProviders, Vendors |
| MetaManagement | Users, DataProviders, AgencyManagement |

| OBJECT | EQUIVALENCE |
|---|---|
| DataManagement | for every 1 data federation |
| DataManagement | for every 10 nonfederated DataProvider |
| CorporateDataManagement | for every 12 Users and 20 DataProviders |
| RepositoryApplications | for up to 100 Users |
| DataAccessApplications | for every 10 DataProviders |
| UATManagement | for every 20 Users |
| InternalApplications | for every 10 Users (internal) |
| InternalUserManagement | for every 20 Users (internal |
| ExternalUserManagement | for every 50 Users (external) |
| NetworkTechnologyManagement | for up to 100 Users+DataProviders |
| DBMSTechnologyManagement | for up to 50 Users and 20 DataProviders |
| GISTechnologyManagement | for up to 50 Users and 20 DataProviders |
| MetaManagement | for every 6 Users (internal) |

In the table, internal users are the number of people required to manage the system. Thus, the process of calculating the resources required must be done iteratively.

The model can now be demonstrated by calculating the human resources required for varying system sizes (in terms of data and users). As an example, let us investigate the resources required to manage the LII if there are 10 DataProviders and 2 federated databases. The following table identifies the number of management objects — which are assigned to people — required to manage the system as a function of the number of users. (See Object Users, following page.)

The table reveals that, with the estimated assignment of effort to responsibility, the management required per user decreases (as we might hope).

This esponsibility groupings which are mapped to a technical architecture and ending with the resources required, in numbers of people, to manage the system.

## Summary

The objective of this paper was to present a model which could aid in the determination of the human resources needed to manage various architectures for geographic information systems. An object-oriented model of the management responsibilities was presented and applied to British Columbia's Land Information Infrastructure. Using the management classes of the model and estimating the effort required to carry out each set of responsibilities, a relationship between the system usage and the management resources was identified.

The model, and the methodology presented here provides a means of identifying management resources prior to system construction and to build the management structure into the system design.

## References

Judge, P. 1988. The Guide to OSI and its Implementation. QED Information Sciences, Inc. Wellesley, MA.

QED. 1989. Strategic and Operational Planning for Information Services. QED Information Sciences, Inc. Wellesley, MA.

Gorman, M. 1991. Database Management Systems. QED Information Sciences, Inc. Wellesley, MA.

Martin, J. Managing the Data-Base Environment. Prentice-Hall, Inc. Englewood Cliffs, NJ.

BC Environment, Lands and Parks. 1991A. Land Information Infrastructure System Architecture. prepared by MacDonald Dettwiler 1991. Victoria, B.C.

BC Environment, Lands and Parks. 1991B. Land Information Infrastructure High Level Design. prepared by MacDonald Dettwiler 1991. Victoria, B.C.

| OBJECT | 10 Users | 50 Users | 200 Users |
|---|---|---|---|
| DataManagement | 2.8 | 2.8 | 2.8 |
| CorporateDataManagement | 0.6 | 1.5 | 6.3 |
| RepositoryApplications | 1.0 | 1.0 | 2.0 |
| DataAccessApplications | 1.0 | 1.0 | 1.0 |
| UATManagementObject | 0.5 | 2.5 | 10.0 |
| InternalApplications | 1.1 | 1.6 | 3.4 |
| InternalUserManagement | 0.6 | 0.8 | 1.7 |
| ExternalUserManagement | 0.2 | 1.0 | 4.0 |
| NetworkTechnology | 1.0 | 1.0 | 2.0 |
| DBMSTechnologyManagement | 1.0 | 1.0 | 3.0 |
| GISTechnologyManagement | 1.0 | 1.0 | 3.0 |
| MetaManagement | 1.8 | 2.5 | 6.5 |
| TOTAL: | 12.6 | 17.7 | 45.7 |
| Management/User | 1.3 | 0.35 | 0.23 |

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# New GIS/LIS education and training initiatives in Australia and southeastern Asia

**Stephen R. Kessell**

Department of Geographic Information Systems
School of Computing
Curtin University of Technology
Perth, Western Australia 6001

## Abstract

*The extraordinary growth in GIS development and applications, especially in the natural resources and environmental fields, in the Australasian region has generated great demand for education and training. During the last few years, Curtin University has become one of the country's largest providers of such GIS instruction. Recent initiatives include:*

> *Introduction of the country's first Bachelor of Science in GIS degree course, with minors offered in a range of subjects including environmental biology and environmental information science;*
>
> *A GIS specialty minor available in the Bachelor of Science in computing science degree;*
>
> *A Bachelor of Science in environmental studies that includes a strong GIS component;*
>
> *A 1-year Graduate Diploma in Geographic and Land Information Systems aimed at retraining mature age students; and Short courses, ranging from 2 days to 6 months, designed specifically for local and overseas mature-age professionals.*

*This paper discusses the rationale and objectives of these widely different degree courses and training programs, and identifies likely future trends and training requirements in the region.*

## Introduction

I am called upon, from time to time, to promote our various Faculty of Science degree programs to high school students, parents and counsellors. One of my favourite themes is "All Scientists are Environmental Scientists". My main point is that environmental science encompasses not only biology and ecology, as important as they may be, but also geology, geomorphology, meteorology, hydrology and lots of other "ologies". Information processing, modelling and especially GIS are fundamental components of modern environmental science and environmental management. But many university degree programs fail to address these needs adequately (Kessell 1990).

In a feature article in *The Australian* newspaper, Gunn and Kennedy (1990) noted that: *The much-chronicled exodus of students from science coursesin favour of degrees such as law, economics and commerce could be on the wane as undergraduates turn "green". It seems the Me Generation is coming to an end, with* *environmentally aware students spurning the yuppie way of assured high-income-leading degrees. Many students seem to be concerned with the future of the world they live in and are determined to have some control over its direction. They are making environmental studies their first choice. A number of universities report a dramatic increase in the numbers of students wishing to do environmental studies at undergraduate level.*

I suspect this is a world-wide, not just as Australian, trend.

There are, however, problems with environmental studies being "flavour of the month". Many prospective students think such a degree will involve subjects such as:

Environmental Studies 101 (syllabus: newspaper recycling, making picket signs, organising protest marches, etc.)

Nature Appreciation 102 (syllabus: "hello trees, hello birds, hello sky")

Upon finding out that such a degree involves a lot of hard science, mathematics and computing, many students will run for cover (Kessell 1990).

I think the real questions are where does the range of analytical tools that we now call "GIS" fit into university degree programs, and are existing degree programs really equipping students for the jobs that await them after graduation. While GIS traditionally has been taught via geography and surveying departments in Australia (AURISA 1989), we have taken a rather different approach at Curtin University.

## GIS at Curtin University

The development of GIS/LIS in Australia has been somewhat different from that in North America, due to:

- the country's large land area (about the same as continental USA) but small population (about 17 million)
- the early development of state land information systems (emphasis on cadastral data)
- individual states, rather than the federal government, being responsible for land management broadly (i.e., each state has its own "national park service")
- dependence on primary production and mining for export income
- proximity to rapidly developing nations in southern Asia.

As a result, "GIS" and "LIS" have developed along different lines. LIS are well-developed in all states, usually in government lands departments; title offices, utilities and transportation agencies also are involved. Tertiary instruction in such LIS is usually provided by academic surveying departments.

In contrast, the development of GIS has been less-centralised and highly concentrated on environmental issues; spatial analysis, modelling, simulation and decision support have driven this development. At the University of New South Wales, it is the Department of Geography that has met research, education and training needs in this "GIS/spatial analysis" field; however, at Curtin University, the initiatives have come from the School of Computing (Kessell and Moore 1989).

Why is GIS centred in the School of Computing? Why not? At the risk of adding to the "what is GIS?" debate, we see it very much as applied computing science. I personally view GIS as an applied science, and think it is about where personal computing was ten or fifteen years ago — growing rapidly and now exploding into many, many different fields.

When the School of Computing was established as a separate entity in 1988 (from the former School of Mathematics and Computing), we had to determine what our Bachelor of Science degree in computing science should include. We took the view that all students must complete a set of "core" courses and must also complete a "specialty" or "minor" applications stream. Thus, in addition to 21 mandatory courses in computing science, mathematics and technical writing, each students must complete another 6 to 8 courses in the applications field of his/her choice. One of the options offered is GIS.

This degree thus produces graduates who are computing scientists firstly, but also have quite intensive training in spatial data bases, computational geometry, spatial analysis, modelling, and graphics; there are also mandatory courses covering basic cartography, geodesy, photogrammetry and remote sensing.

While this degree program was being implemented in 1988, a quite different GIS/LIS education need surfaced: mature age professionals, in disciplines ranging from forestry to town planning, were also seeking training/re-training in GIS. Our response was the establishment of a Graduate Diploma in Geographic and Land Information Systems, taught jointly by the Schools of Computing and Surveying. While the program may be completed in one year's full time study, most students are employed full-time and study part-time, completing the diploma over a two or three year period (all mandatory classes are available in the evening to accommodate this). More recently, the graduate diploma has attracted recent graduates from environmental science fields; they have completed the diploma before seeking full-time employment.

**Table 1. Course structure for the Bachelor of Science in GIS degree.**

- **Year 1**
  - First Semester
    - Computer Programming 151
    - Computer Equipment 151
    - Mathematics 101
    - Mathematical Statistics 101
    - English for Technical Communication 100
  - Second Semester
    - Computer Programming 152
    - Operating Systems 152
    - Mathematics 102
    - Applied Statistics 102
    - Information Technology and Society 152
- **Year 2**
  - First Semester
    - Scientific Computing 251
    - Geographic Information Systems 251
    - Computational Geometry 201
    - units from minor stream
  - Second Semester
    - Information Storage and Retrieval 252
    - Geodesy 202
    - Environmental Information Analysis 252
    - units from minor stream
- **Year 3**
  - First Semester
    - Technical Writing and Project Preparation 391
    - GIS Applications 351
    - Photogrammetry and Remote Sensing 301
    - units from minor stream
  - Second Semester
    - Spatial Analysis, Modelling and Data Structures 352
    - GIS Project 394
    - units from minor stream / electives

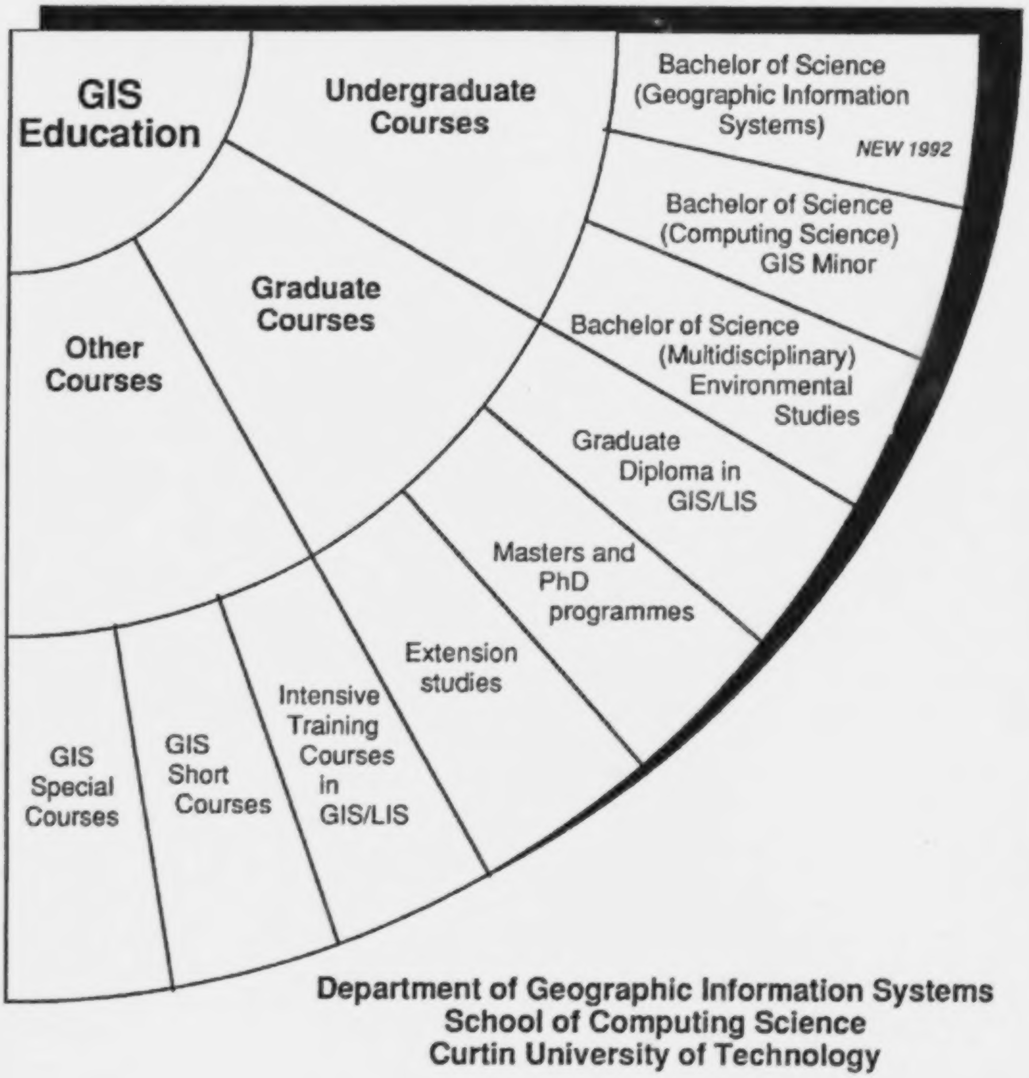

GIS Education
Undergraduate Courses
Graduate Courses
Other Courses
Bachelor of Science (Geographic Information Systems)
NEW 1992
Bachelor of Science (Computing Science) GIS Minor
Bachelor of Science (Multidisciplinary) Environmental Studies
Graduate Diploma in GIS/LIS
Masters and PhD programmes
Extension studies
Intensive Training Courses in GIS/LIS
GIS Short Courses
GIS Special Courses
Department of Geographic Information Systems
School of Computing Science
Curtin University of Technology

Figure 1. Relationship among the various GIS education and training programs offered.

Thus there is a sharp contrast between the students learning GIS as part of their undergraduate computing science degree, and those attracted to the graduate diploma. At the risk of over-simplification, I will characterise the former as students who will *program* the next generation of GIS systems and the latter as those who will be intelligent *users* of such systems.

## Broader needs

Despite the immediate success of both GIS programs, it became apparent very rapidly that we were meeting GIS education and training demands only in part. As a result, three other major initiatives have been taken over the past couple years.

Firstly, a new Bachelor of Science in environmental studies (really environmental information analysis) has been developed via the university's multidisciplinary B S degree. Students must select both a "major" (from environmental biology, environmental chemistry, physics or GIS) and complete a "minor" in two of the other disciplines. This structure was selected specifically to ensure graduates have a strong computational / GIS background. Many of the individual courses are also taken as electives from students completing other degrees (in biology, town planning, geography, etc.).

Secondly, a range of short courses was developed; we currently offer not only the traditional 1, 2, 3, or 5 day course, but also an intensive 6-month course aimed particularly at overseas professionals. The 6-month course is tailored for each particular group, but invariable includes GIS theory and practice, microcomputing, programming and spatial analysis. "Background" classes reviewing basic mathematics, statistics, cartography, etc. often are included, as well as intensive English as a second language instruction. Vendor-specific training completes the course.

Finally, we have just taken the bolder step of establishing a Department of GIS within the School of Computing, and offering a Bachelor of Science in GIS degree (possibly the first in the world, I am told). Why a GIS degree? We perceive, in Australia and probably elsewhere, a need for professionals educated intensively the various aspects of GIS, all the way from the data base and programming side to the applications areas. We have structured the degree so that all students must complete a set of mandatory courses (this provides the theory and numeracy); they must also select a minor specialty (applications) area. Likely minors include environmental studies, town planning, geography, cartography, image processing and computing science. The syllabus for this 3-year degree is shown in Table 1. Figure 1 attempts to put the range of our GIS offerings into some perspective.

## Now what ?

Perhaps we should pause and catch our breath. With a solid range of GIS undergraduate degree / graduate diploma training courses in place, the next emphasis will be the development of the graduate and research program. Just as we observed when the new computing science school and degree were established in 1988, it takes a few years to produce a crop of graduates who are suitably prepered for graduate study. While we accept some graduates from other universities into masters and doctorate programs, I do not anticipate a full-fledged graduate school until the mid 90's. In the meantime, it appears that the demand for graduates will remain strong (this demand has survived the current economic climate very well indeed) and that the demand for a range of different training programs and courses will continue to grow.

## Thanks

This paper was written while on vacation in the wilds of Maine. Many thanks to Trillium Computer Systems, of Norway, Maine, for providing access to word processing facilities.

## Literature cited

AURISA. (1989) *Report of the National Working Party on Education and Research in Geographic and Land Information Systems.* Australasian Urban and Regional Information Systems Assoc., Eastwood, N.S.W.

Gunn, M. and F. Kennedy. (1990) The rise and rise of the Tree Generation. *The Australian Newspaper.* 21st February.

Kessell, S. R. (1990) *The role of GIS in a university environmental studies course.* Proc. AURISA(18) (Australasian Urban and Regional Information Systems Assoc. Conf.): 27-32.

Kessell, S. R. and D. Moore (1989) *Meeting diverse needs in GIS training and education.* Australasian Urban and Regional Information Systems Assoc., Proc. of URPIS 17 (Perth): 231-244.

# Reserve and poster papers

*The following papers will either be presented at GIS'92 as reserve papers or as posters.*

**GIS management of felled and bucked log inventories susceptible to Ambrosia beetle attack**

Rick Hawe et al., Phero Teck Inc., Delta, BC, Canada

**Recent developments in spatial analysis of fisheries acoustic data**

R. Kieser et al., Pacific Biological Station, Nanaimo, BC, Canada

**The next generation of digital spatial technologies: error utilization, thematic map and remote sensing integration, spatial operators and data structures**

Kim Lowell et al., Université Laval, Sainte Foy, PQ, Canada

**Spatial considerations for forest canopy modelling and biomass estimation based on remotely sensed data**

Geoffrey Edwards et al., Université Laval, Sainte Foy, PQ, Canada

**Regeneration stocking derived from high resolution airborne imagery via texture measurements**

Geoffrey Edwards, Université Laval, Sainte Foy, PQ, Canada

**An expert system for helping forest managers choose among data sources of remotely sensed imagery**

Geoffrey Edwards et al., Université Laval, Sainte Foy, PQ, Canada

**A geographic information system for strategic and operational monitoring of forest activity: a conceptual approach**

Riadika Mastra et al., PT Mapindo Parama, Jakarta, Indonesia

**Eco-camp: and environmental education experience**

Timothy Day et al., California State Polytechnic University, Pomona, CA, USA

GIS'92

Symposium

Vancouver
British Columbia
February 1992

# GIS management of felled and bucked log inventories susceptible to ambrosia beetle attack

**Rick Hawe**

Pest Management Forester
Phero Teck Inc
7572 Progress Way
Delta, B.C. V4G 1E9

**Dr. John McLean**

UBC Faculty of Forestry

## Abstract

*This paper details the development of an inventory management system designed to reduce the damage inflicted by* Trypodendron lineatum, *the most common ambrosia beetle on the B.C. coast. The system was developed for Canadian Forest Products Ltd. operating in the Nimpkish valley on Northern Vancouver Island. This inventory management application utilizes Terrasoft, a PC based Geographic Information System. The application was designed to track the progression of felling and yarding operations in each setting.*

*Using this information, the system alerts production personnel to those areas which are at risk to ambrosia beetle attack. The areas identified as being at risk in each setting are displayed graphically and are color coded according to the value of the timber. With this tool production personnel will be able to priorize logging activities and reduce the losses associated with ambrosia beetles.*

## Introduction

*Project background* This paper is a summary of the methodologies and findings of previous research presented in an undergraduate thesis. The original thesis was prepared for the University of B.C. in partial fulfillment of the requirements for a Degree of Bachelor of Science in Forestry. The original research was done in conjunction with the Englewood logging division of Canadian Forest Products Ltd., in Woss B.C. A total of Eight logging settings were included in the study. This paper outlines the ambrosia beetle problem, describes the methods used in the study, and presents the results for setting KT-126, one of the eight logging sites.

### Life history of ambrosia beetles

Ambrosia beetles are small, cylindrical, elongated insects of the beetle families Platypodidae and Scolytidae. These insects bore directly into the wood of weakened or recently killed trees, fresh logs, and occasionally green lumber (Johnson, 1958). The striped ambrosia beetle, *Trypodendron lineatum*, is the most prevalent species on the Coast (Shore, et al., 1987).

*T. lineatum* overwinter in the forest, seeking protection in the duff on the forest floor. In spring when their overwintering site has warmed up sufficiently and the air temperature is above 16 degrees centigrade, the beetles emerge and fly to attack new logs (Nijholt, 1978).

### Ambrosia beetles and logging

B.C. is plagued with chronic infestations of ambrosia beetles due to the wealth of suitable breeding material. Logging activities and natural causes, such as disease, insects, and blow down during winter storms provide suitable habitat for developing broods (McLean and Stokkink, 1988).

Control efforts have been focused mainly on those areas where large inventories of timber are stored; dry land sorts, mill yards, and booming grounds. Control techniques in these areas consist of mass trapping the ambrosia beetles using pheromones and specially designed traps.

Studies during the last decade have shown that many ambrosia beetle infestations originate in the forest (McLean and Stokkink, 1988). Unfortunately mass trapping techniques are presently considered ineffective and uneconomical in forest settings where the susceptible timber is spread over a large area. Consequently the need to improve management of log inventories in active settings was recognized, in order to reduce the degree of infestation which occurs before the logs reach the storage areas (McLean and Salom, 1989).

This report proposes such a system for managing log inventories in active settings in an attempt to reduce the incidence of attack. The system utilizes Terrasoft, a PC based Geographic Information System (GIS).

### The logging process

To prepare timber for yarding, settings have to befelled well in advance. Areas slated for logging in the winter or spring are felled in the late summer or fall. Settings scheduled to be yarded in the summer or fall are felled in the spring and early summer.

*T. lineatum* will attack timber which has been felled a minimum of 2 months (Gray & Borden, 1985). Depending on where the attack occurs, the implications for subsequent attacks and population buildups can be significant. Logs supporting a developing brood, will house the beetles for 8 -10 weeks. At the end of this period the beetles will emerge and search for overwintering sites in preparation for next year's spring attack. If the logs were originally attacked in the setting and end up at the dryland sort when the brood emerges then the beetle population will increase at that site. Effective inventory management and timing of operations are very important.

### Economic aspects of ambrosia beetles

Ambrosia beetle damage significantly reduces the value of products manufactured from logs in Western North America. These effects result in reduced profits from lower grades of lumber and plywood, and lost export opportunities.The more lucrative offshore markets often place severe restrictions on beetle damaged wood products. In addition to the losses resulting from lumber degrade, other processing costs reduce the net value of lumber bearing ambrosia beetle damage. These costs are incurred through declines in lumber recovery, increased handling time, increased slabbing, and manufacturing.

Industry estimates suggest that in uncontrolled situations 1 out of every 3 logs being processed on the coast will bear some degree of ambrosia beetle damage. Experts have estimated that this damage costs the forest industry in excess of $200 million annually. A single manufacturing facility processing an average of 120 000 Mfbm per year stands to lose $1.8 to $2 million annually (Anon., 1990).

Logging companies view the whole logging process as a series of cumulative costs. The efficiency of each step, from falling to delivering the logs at the mill, is judged by the cost per cubic meter. No assessment of value preservation is currently included in determining the efficiency of the operation.

### CANFOR's operations

Annual loss to Canadian Forest Products Ltd. (CANFOR) from ambrosia beetle attacked logs originating on the Englewood Logging Division (TFL #37) is estimated at $2 million (1988 figures). According to Phero Tech trapping records, *Trypodendron lineatum* accounts for 95% of the resident ambrosia beetle population (Wilson, 1988).

The Englewood logging division cuts approx. 1.3 million cubic meters of wood annually, and supplies 65% of the timber processed at CANFOR's Eburne sawmill. The mill's entire design is predicated on acquiring high grade logs and producing high grade lumber. There is a high leverage in profit or loss at Eburne relating to log quality (Wilson, 1988).

The objective of this paper is to demonstrate that using Terrasoft as a tool to aid inventory management, CANFOR can help maintain this flow of high quality timber to the mill. The inventory management system proposed here keeps track of the felled and bucked timber, and the parameters associated with it. The system was designed to inform production personnel of those areas which are considered to be at risk to attack by *T. lineatum* during the spring flight period. This information can be accessed at any time and will aid in the planning and management of logging operations.

## Methods

### GIS procedure

At the present time Canadian Forest Products maintainsactive setting files for each Cutting permit (CP) in TFL #37 which covers a total of 195 000 hectares.

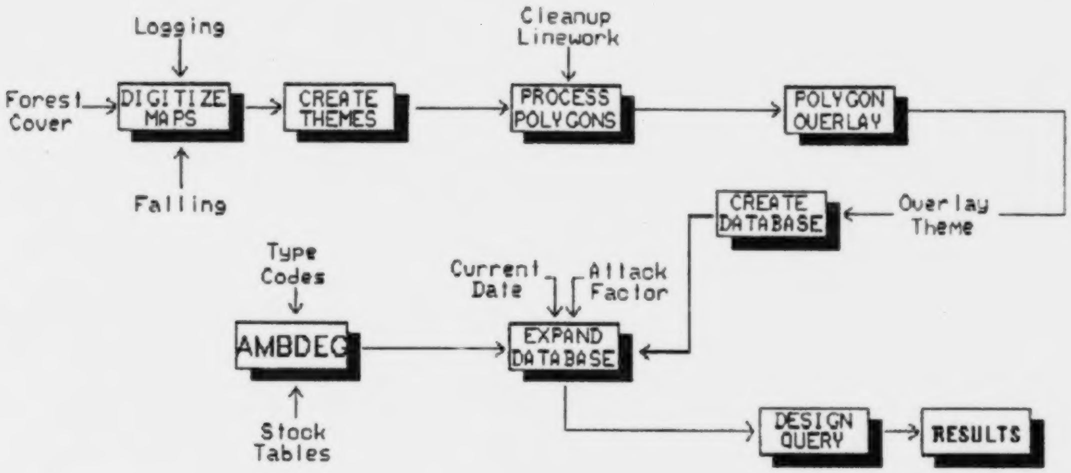


**Figure 1:** *The GIS Process*

My objective was to take the existing information fromthe setting file maps and enter it into Terrasoft. This information was then used to form a database file containing data relating to felling, logging, and other parameters. The intention was to analyze this data through Terrasoft and identify the areas and values at risk to attack by *T. lineatum*. Such a felled and bucked timber management system could easily be expanded to cover all active settings within the TFL, and provide methods for updating the records on a monthly basis. The following is a detailed description of the procedures involved in developing this system.

### The GIS process

The flowchart (Figure 1) outlines the steps involved in developing a workable system with the desired results. The initial step involved digitizing the block boundary, forest cover, falling, and logging data from the individual setting maps. The linework for each of these attributes was placed on a separate level. In addition, labels identifying the forest cover type, month of falling, and month of logging were also placed on separate levels.

The digitized linework was cleaned up to ensure the line segments joined properly, and themes were created for forest cover, falling, and logging. For example; the forest cover theme required the 3 levels which contained information about the forest cover: the block boundary linework level, the forest cover linework level, and the forest cover label level.

Once the individual themes were created the linework was processed to form polygons. With each theme processed into polygons containing their associated labels, the next step was to overlay the three themes. The overlay procedure creates new child polygons which are formed from the intersection of the linework from the parent polygons in the three themes. Figure 2 a-c shows the linework and labels associated with the three themes that were overlayed to form the child polygons in setting KX-23.

The new child polygons are automatically assigned a linknumber, which was also stored on a graphic level. Using the links as a base, Terrasoft forms a relational database file. This database file contains the information from the three themes in addition to other information which is relevant to each of the new polygons. The following table shows the initial database structure created by Terrasoft.

As the overlay procedure continues each new polygon is tagged with the labels from the parent polygon themes. As a result each child polygon has a database record which lists the forest cover type it is in, when it was felled, and when it was logged. Other polygon attributes pertaining to the size and position of the polygons are calculated and recorded in the database.

### Expanding the database (part 1)

***Forest cover type codes:*** When Canadian Forest Products re-inventoried TFL #37 107 different forest types were identified based on the inventory plot data. The different types were identified by season of accessibility, species, age, and height. Each of the 107 types had a stock table associated with it. The stock tables show a breakdown of the volume in each forest type by species and diameter class (Table 2).

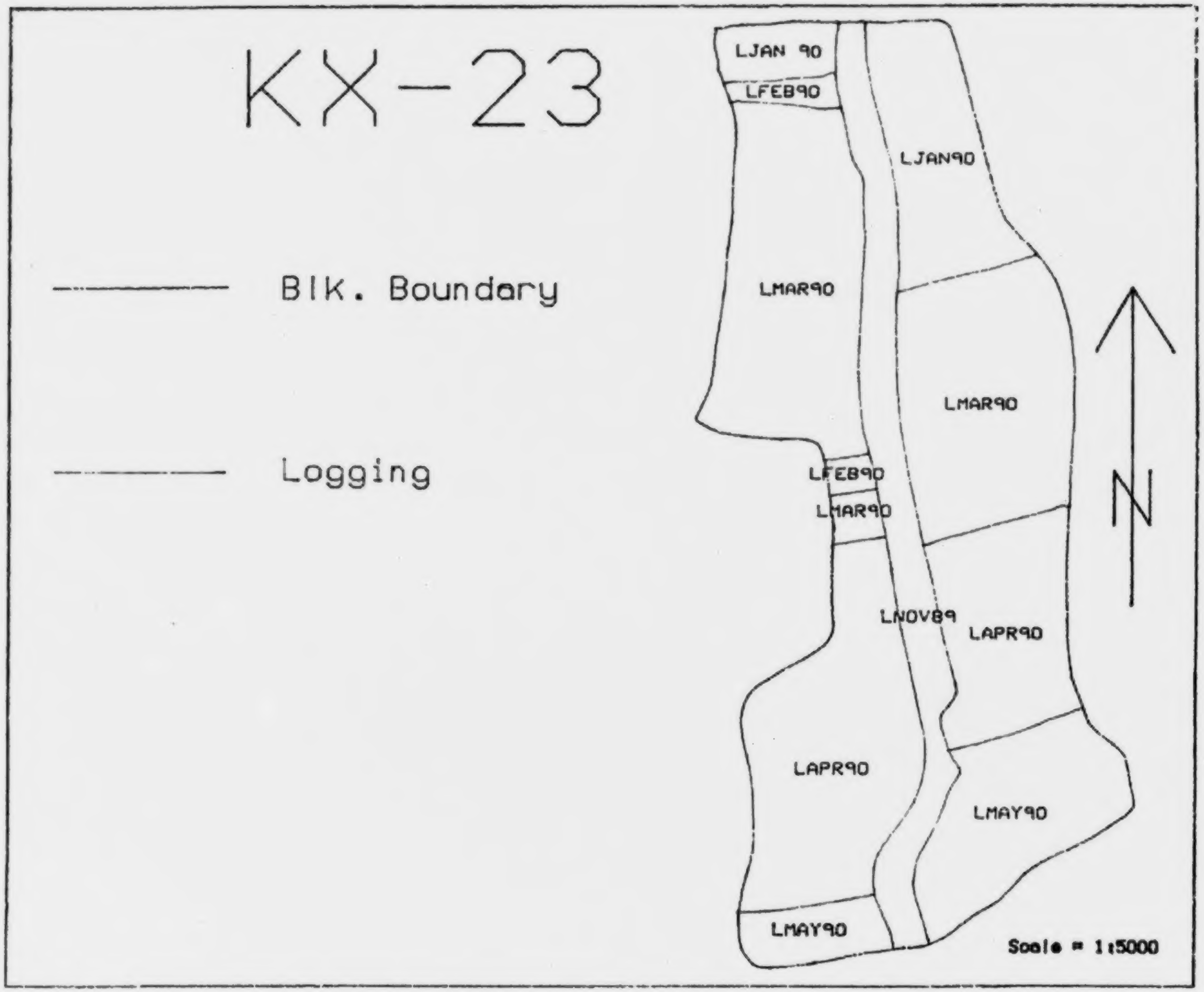

KX-23
Blk. Boundary
Logging
LJAN 90
LFEB90
LJAN90
LMAR90
LMAR90
LFEB90
LMAR90
LNOV89
LAPR90
LAPR90
LMAY90
LMAY90
N
Scale = 1:5000

Figure 2c: *Logging Theme*

## Table 1: Initial database structure

| POLY_LABEL | POLY_AREA | POLY_X | POLY_Y | POLY_STAT | GIS_LINK | POLY_PERIM | FC | LOGGING | FALLING |
|---|---|---|---|---|---|---|---|---|---|

POLY_LABEL = Polygon label (Numerical & same as GIS_LINK)
POLY_AREA = Area of the individual polygons
POLY_X = X co-ordinate position of the polygon
POLY_Y = Y co-ordinate position of the polygon
POLY_STAT = Status of the polygon, tells whether it is properly linked or not
GIS_LINK = Numerical link associating database records to graphical polygons
POLY_PERIM = Perimeter of the individual polygons
FC = The forest cover type that the polygon is in. (Tagged from forest cover theme)
LOGGING = Month of logging (Polygon is tagged from original logging theme)
FALLING = Month of falling (Polygon is tagged from original falling theme)

## Table 2: Stock table for timber type 1

| STOCK TYPE | Dbh. (cm) | D Fir (cu. m) | Cedar (cu. m) | Hemlock (cu. m) | Balsam (cu. m) | Spruce (cu. m) | Y Cedar (cu. m) | M Hem (cu. m) | W Pine (cu. m) | L Pine (cu. m) | Alder (cu. m) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 20 | | | | 19 | | | | | | | |
| 1 | 25 | | | 12.3 | | | | | | | | |
| 1 | 30 | | | | | | | | | | | |
| 1 | 35 | | | 16.6 | | | | | | | | |
| 1 | 40 | | 10.9 | | | | | | | | | |
| 1 | 45 | 15 | 10.5 | 18.2 | | | | | | | | |
| 1 | 50 | | | | | | | | | | | |
| 1 | 55 | | 12.4 | 19.1 | | | | | | | | |
| 1 | 60 | | | | | | | | | | | |
| 1 | 65 | 7.3 | | | | | | | | | | |
| 1 | 70 | | 12.3 | 21.2 | | | | | | | | |
| 1 | 75 | 18.8 | 19.1 | | | | | | | | | |
| 1 | 80 | | | | 12.2 | | | | | | | |
| 1 | 85 | | | | | | | | | | | |
| 1 | 90 | | | | | | | | | | | |
| 1 | 95 | | | 23.7 | | | | | | | | |
| 1 | 100 | | | | | | | | | | | |
| 1 | 105 | | 20.7 | | | | | | | | | |
| 1 | 110 | | | | | | | | | | | |
| 1 | 115 | | 34.5 | | | | | | | | | |
| 1 | 120 | 35.2 | | | | | | | | | | |
| 1 | 125 | | 25.7 | | | | | | | | | |
| 1 | 130 | | 11.6 | | | | | | | | | |
| 1 | 135 | 23.5 | | | | | | | | | | |
| 1 | 140 | | 12.3 | | | | | | | | | |
| 1 | 145 | | 15.3 | | | | | | | | | |
| 1 | 150 | | | | | | | | | | | |
| 1 | 155 | | 12.2 | | | | | | | | | |
| 1 | 160 | | | | | | | | | | | |
| 1 | 165 | | | | | | | | | | | |
| 1 | 170 | | 24.9 | | | | | | | | | |
| 1 | 175 | | 12.4 | | | | | | | | | |
| 1 | 180 | | 15.9 | | | | | | | | | |
| 1 | 185 | | | | | | | | | | | |
| 1 | 190 | | | | | | | | | | | |
| 1 | 195 | | | | | | | | | | | |
| 1 | 200 | | 11.5 | | | | | | | | | |
| Totals | | 99.8 | 262.2 | 111.1 | 31.2 | 0 | 0 | 0 | 0 | 0 | 0 | 504.3 |

## Table 3: Summary of type codes for forest cover labels in setting KT-126

| POLY_LABEL | SPECIES | AGE | HT | TYPE |
|---|---|---|---|---|
| KT126.1 | HBCy | D | 5 | 99 |
| KT126.10 | CyHB | D | 3 | 79 |
| KT126.11 | BH | D | 4 | 77 |
| KT126.12 | BH(F) | D | 4 | 77 |
| KT126.13 | BH | D | 4 | 77 |
| KT126.14 | BH | D | 3 | 77 |
| KT126.15 | CyHB | D | 3 | 79 |
| KT126.16 | HBCy | D | 3 | 83 |
| KT126.17 | CHBCy | D | 3 | 73 |
| KT126.18 | CyCH | D | 4 | 96 |
| KT126.19 | HBCy | D | 3 | 75 |
| KT126.2 | HB | D | 4 | 99 |
| KT126.20 | HCyB | D | 3 | 72 |
| KT126.21 | CyHB(F) | D | 3 | 74 |
| KT126.22 | CyHB | D | 3 | 74 |
| KT126.3 | HBCy | D | 3 | 83 |
| KT126.4 | CyHB | D | 2 | 67 |
| KT126.5 | HB | D | 3 | 99 |
| KT126.6 | HBCy | D | 3 | 83 |
| KT126.7 | HB | D | 3 | 99 |
| KT126.8 | BH | D | 4 | 77 |
| KT126.9 | CyHB | D | 3 | 79 |

Table 2 shows the breakdown of the volume for forest type #1, similar tables exist for the remaining 106 stock types. The database that was initially created by Terrasoft when the polygon overlay procedure was executed contained the records for forest cover. These records were based on the forest cover labels of the parent polygons.

The parent polygons were labeled according to the setting, and polygon number. For example KT126.4 referred to forest cover parent polygon #4 in setting KT-126. Each of these labels had a species, age, and height class associated with it which identified the actual forest cover. Knowing the species, age, and height class, as well as the season I was able to determine the correct type code from the list of 107. Table 3 shows all the type codes associated with the polygon forest cover labels for setting KT-126.

The main database containing the records for the child polygons was updated with the new inventory type code. This was achieved by creating a new field within this database. The type code information was then used to access the individual stock tables for each of the timber types. These stock tables contain the information that is required to calculate the amount of degrade that would result from ambrosia beetle attack. The breakdown of the volume by species and diameter class for each of the type codes is utilized by a computer degrade model called AMBDEG. This degrade model was developed by Dr. John A. McLean, a professor of forest entomology at the University of B.C.

### AMBDEG

This model was developed to evaluate the degrade caused by ambrosia beetles based on the volume of wood that is attacked. This volume is usually associated with the highest value clear wood on the outer portions of the logs. The degrade value is calculated as the difference between the value of clear lumber and the value of lumber with beetle pinholes. These values are computed using the current Lumber Recovery Factors (LRF) for each type of log (McLean, 1990).

***AMBDEG parameters:*** Within the AMBDEG program the following three parameters had to be set: LRF, Log Degrade Value, and depth of beetle penetration. The LRF refers to the amount of cut lumber that can be processed from an un-sawn log at the mill. The LRF for this study wasset at 275 fbm/cu. m. This value is easily changed within the AMBDEG program, and should be verified with the sawmill involved before calculations are initiated (McLean, 1991).

The Log Degrade Value refers to the differential value between clear lumber and that same lumber with pinholes from the grade of log being evaluated (McLean,

## Table 4: AMBDEG summary for hemlock

-------- Log Summary --------

Lumber recovery factor = 275.00
Log degrade value = 500.00

| Log ID | Spec. | Top Dia. (cm) | Bottom Dia. (cm) | Length (m) | Volume (m3) | Attack Depth (cm) | Prop. Degrade | Degrade Vol. (m3) | Degrade Loss ($) | Degrade Loss ($/m3) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | HB | 20 | 20 | 10 | 0.31 | 8 | 0 | 0 | $0.00 | $0.00 |
| 2 | HB | 25 | 25 | 10 | 0.49 | 8 | 0 | 0 | $0.00 | $0.00 |
| 3 | HB | 30 | 30 | 10 | 0.71 | 8 | 0 | 0 | $0.00 | $0.00 |
| 4 | HB | 35 | 35 | 10 | 0.96 | 8 | 0.028 | 0.027 | $3.71 | $3.86 |
| 5 | HB | 40 | 40 | 10 | 1.26 | 8 | 0.063 | 0.079 | $10.89 | $8.64 |
| 6 | HB | 45 | 45 | 10 | 1.59 | 8 | 0.098 | 0.156 | $21.43 | $13.48 |
| 7 | HB | 50 | 50 | 10 | 1.96 | 8 | 0.133 | 0.261 | $35.91 | $18.32 |
| 8 | HB | 55 | 55 | 10 | 2.38 | 8 | 0.168 | 0.399 | $54.88 | $23.06 |
| 9 | HB | 60 | 60 | 10 | 2.83 | 8 | 0.203 | 0.574 | $78.92 | $27.89 |
| 10 | HB | 65 | 65 | 10 | 3.32 | 8 | 0.238 | 0.79 | $108.59 | $32.71 |
| 11 | HB | 70 | 70 | 10 | 3.85 | 8 | 0.273 | 1.051 | $144.46 | $37.52 |
| 12 | HB | 75 | 75 | 10 | 4.42 | 8 | 0.308 | 1.361 | $187.10 | $42.33 |
| 13 | HB | 80 | 80 | 10 | 5.03 | 8 | 0.343 | 1.724 | $237.07 | $47.13 |
| 14 | HB | 65 | 85 | 10 | 5.67 | 8 | 0.341 | 1.935 | $266.10 | $46.93 |
| 15 | HB | 90 | 90 | 10 | 6.36 | 8 | 0.324 | 2.061 | $283.37 | $44.56 |
| 16 | HB | 95 | 95 | 10 | 7.09 | 8 | 0.308 | 2.187 | $300.65 | $42.40 |
| 17 | HB | 100 | 100 | 10 | 7.85 | 8 | 0.294 | 2.312 | $317.93 | $40.50 |
| 18 | HB | 105 | 105 | 10 | 8.66 | 8 | 0.282 | 2.438 | $335.21 | $38.71 |
| 19 | HB | 110 | 110 | 10 | 9.5 | 8 | 0.27 | 2.564 | $352.49 | $37.10 |
| 20 | HB | 115 | 115 | 10 | 10.39 | 8 | 0.259 | 2.689 | $369.77 | $35.59 |
| 21 | HB | 120 | 120 | 10 | 11.31 | 8 | 0.249 | 2.815 | $387.05 | $34.22 |
| 22 | HB | 125 | 125 | 10 | 12.27 | 8 | 0.24 | 2.941 | $404.33 | $32.95 |
| 23 | HB | 130 | 130 | 10 | 13.27 | 8 | 0.231 | 3.066 | $421.61 | $31.77 |
| 24 | HB | 135 | 135 | 10 | 14.31 | 8 | 0.223 | 3.192 | $438.88 | $30.67 |
| 25 | HB | 140 | 140 | 10 | 15.39 | 8 | 0.216 | 3.318 | $456.16 | $29.64 |
| 26 | HB | 145 | 145 | 10 | 16.51 | 8 | 0.209 | 3.443 | $473.44 | $28.68 |
| 27 | HB | 150 | 150 | 10 | 17.67 | 8 | 0.202 | 3.569 | $490.72 | $27.77 |
| 28 | HB | 155 | 155 | 10 | 18.87 | 8 | 0.196 | 3.695 | $508.00 | $26.92 |
| 29 | HB | 160 | 160 | 10 | 20.11 | 8 | 0.19 | 3.82 | $525.28 | $26.12 |
| 30 | HB | 165 | 165 | 10 | 21.38 | 8 | 0.185 | 3.946 | $542.56 | $25.38 |
| 31 | HB | 170 | 170 | 10 | 22.7 | 8 | 0.179 | 4.072 | $559.84 | $24.66 |
| 32 | HB | 175 | 175 | 10 | 24.05 | 8 | 0.174 | 4.197 | $577.11 | $24.00 |
| 33 | HB | 180 | 180 | 10 | 25.45 | 8 | 0.17 | 4.323 | $594.39 | $23.36 |
| 34 | HB | 185 | 185 | 10 | 26.88 | 8 | 0.165 | 4.449 | $611.67 | $22.76 |
| 35 | HB | 190 | 190 | 10 | 28.35 | 8 | 0.161 | 4.574 | $628.95 | $22.19 |
| 36 | HB | 195 | 195 | 10 | 29.86 | 8 | 0.157 | 4.7 | $646.23 | $21.64 |
| 37 | HB | 200 | 200 | 10 | 31.42 | 8 | 0.154 | 4.825 | $663.51 | $21.12 |

1990). In this particular study a Log Degrade Value of $500/mfbm was chosen.

The depth of beetle penetration affects how much the outer portion of the log is degraded by the beetle infestation. This value is dependent on the species of conifer log because ambrosia beetles tend to damage some species more than others. The attack depth was set at 8 cm. for Hemlock and Balsam, and 6 cm. for Fir, Yellow Cedar, and Spruce.

***AMBDEG calculations:*** With these initial parameters set I calculated the degrade loss for standard 10 m. logs ranging from 20 cm in diameter to 200 cm in diameter, in 5 cm increments. The degrade loss for this range of logs was calculated for each species susceptible to ambrosia beetle attack. This included Douglas Fir, Balsam Fir, Spruce, Yellow Cedar, and Hemlock. The degrade calculations were performed with a degrade value of $500/mfbm. Table 4 shows the report generated by AMBDEG for Hemlock.

The summary table generated by AMBDEG gives the volume of a 10 meter log in the different diameter classes. It also gives the total ambrosia degrade loss for each log at the declared LRF, Degrade Value, and Depth of Penetration. With this information one can then calculate the amount of degrade per cubic meter

| STOCK | DBH | VOLUME BY SPECIES & DIAMETER CLASS (cu.m. / ha.) | | | | | | | D Fir Degrade | Hemlock Degrade | Balsam Degrade | Spruce Degrade | Y Cedar Degrade | D Fir Deg. Loss | Hemlock Deg. Loss | Balsam Deg. Loss | Spruce Deg. Loss | Y Cedar Deg. Los | Totals |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TYPE | (cm) | D Fir | Cedar | Hemlock | Balsam | Spruce | Y Cedar | | ($/m3) | ($/m3) | ($/m3) | ($/m3) | ($/m3) | ($/ha.) | ($/ha.) | ($/ha.) | ($/ha.) | ($/ha.) | ($/ha.) |
| 1 | 20 | | | | 19 | | | | $0.00 | $0.00 | $2.23 | $0.35 | $26.98 | $0.00 | $0.00 | $42.29 | $0.00 | $0.00 | $42.29 |
| 1 | 25 | | | 12.3 | | | | | $0.00 | $0.00 | $3.51 | $4.79 | $32.08 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 30 | | | | | | | | $0.00 | $0.00 | $4.79 | $9.15 | $37.25 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 35 | | | 16.6 | | | | | $6.61 | $3.86 | $6.13 | $13.65 | $42.90 | $0.00 | $64.07 | $0.00 | $0.00 | $0.00 | $64.07 |
| 1 | 40 | | 10.9 | | | | | | $14.40 | $8.64 | $7.40 | $17.99 | $48.06 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 45 | 15 | 10.5 | 18.2 | | | | | $22.28 | $13.48 | $8.73 | $22.47 | $53.60 | $334.20 | $245.30 | $0.00 | $0.00 | $0.00 | $579.50 |
| 1 | 50 | | | | | | | | $30.17 | $18.32 | $10.06 | $26.93 | $58.18 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 55 | | 12.4 | 19.1 | | | | | $37.88 | $23.06 | $11.32 | $31.25 | $53.36 | $0.00 | $440.42 | $0.00 | $0.00 | $0.00 | $440.42 |
| 1 | 60 | | | | | | | | $45.75 | $27.89 | $12.64 | $35.69 | $49.46 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 65 | 7.3 | | | | | | | $46.06 | $32.71 | $13.95 | $40.12 | $46.06 | $336.23 | $0.00 | $0.00 | $0.00 | $0.00 | $336.23 |
| 1 | 70 | | 12.3 | 21.2 | | | | | $43.08 | $37.52 | $15.26 | $43.08 | $43.08 | $0.00 | $795.47 | $0.00 | $0.00 | $0.00 | $795.47 |
| 1 | 75 | 18.8 | 19.1 | | | | | | $40.46 | $42.33 | $16.56 | $40.46 | $40.46 | $760.66 | $0.00 | $0.00 | $0.00 | $0.00 | $760.66 |
| 1 | 80 | | | | 12.2 | | | | $38.13 | $47.13 | $17.86 | $38.13 | $38.13 | $0.00 | $0.00 | $217.93 | $0.00 | $0.00 | $217.93 |
| 1 | 85 | | | | | | | | $36.11 | $46.93 | $19.20 | $36.11 | $36.11 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 90 | | | | | | | | $34.23 | $44.56 | $20.49 | $34.23 | $34.23 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 95 | | | 23.7 | | | | | $32.53 | $42.40 | $21.79 | $32.53 | $32.53 | $0.00 | $1,004.99 | $0.00 | $0.00 | $0.00 | $1,004.99 |
| 1 | 100 | | | | | | | | $31.04 | $40.50 | $23.11 | $31.04 | $31.04 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 105 | | 20.7 | | | | | | $29.63 | $38.71 | $24.40 | $29.63 | $29.63 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 110 | | | | | | | | $28.37 | $37.10 | $25.72 | $28.37 | $28.37 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 115 | | 34.5 | | | | | | $27.19 | $35.59 | $27.01 | $27.19 | $27.19 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 120 | 35.2 | | | | | | | $26.12 | $34.22 | $28.32 | $26.12 | $26.12 | $919.57 | $0.00 | $0.00 | $0.00 | $0.00 | $919.57 |
| 1 | 125 | | 25.7 | | | | | | $25.14 | $32.95 | $29.64 | $25.14 | $25.14 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 130 | | 11.6 | | | | | | $24.22 | $31.77 | $30.94 | $24.22 | $24.22 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 135 | 23.5 | | | | | | | $23.36 | $30.67 | $30.67 | $23.36 | $23.36 | $549.06 | $0.00 | $0.00 | $0.00 | $0.00 | $549.06 |
| 1 | 140 | | 12.3 | | | | | | $22.57 | $29.64 | $29.64 | $22.57 | $22.57 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 145 | | 15.3 | | | | | | $21.82 | $28.68 | $28.68 | $21.82 | $21.82 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 150 | | | | | | | | $21.12 | $27.77 | $27.77 | $21.12 | $21.12 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 155 | | 12.2 | | | | | | $20.47 | $26.92 | $26.92 | $20.47 | $20.47 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 160 | | | | | | | | $19.85 | $26.12 | $26.12 | $19.85 | $19.85 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 165 | | | | | | | | $19.28 | $25.38 | $25.38 | $19.28 | $19.28 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 170 | | 24.9 | | | | | | $18.73 | $24.66 | $24.66 | $18.73 | $18.73 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 175 | | 12.4 | | | | | | $18.21 | $24.00 | $24.00 | $18.21 | $18.21 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 180 | | 15.9 | | | | | | $17.72 | $23.36 | $23.36 | $17.72 | $17.72 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 185 | | | | | | | | $17.26 | $22.76 | $22.76 | $17.26 | $17.26 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 190 | | | | | | | | $16.82 | $22.19 | $22.19 | $16.82 | $16.82 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 195 | | | | | | | | $16.41 | $21.64 | $21.64 | $16.41 | $16.41 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| 1 | 200 | | 11.5 | | | | | | $16.00 | $21.12 | $21.12 | $16.00 | $16.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 | $0.00 |
| Total | | 99.8 | 262.2 | 111.1 | 31.2 | 0 | 0 | 504.3 | | | | | | $2,899.72 | $2,550.25 | $260.22 | $0.00 | $0.00 | $5,710.19 |

Table 5: Calculation of degrade losses for stock type 1 @ log degrade value of $500/cu. m.

of wood which is shown in the final column of the summary table. These values were extracted from the tables and formed the basis of the calculations to determine the total degrade for each timber type.

## Calculation of total degrade for each timber type

The calculation of the degrade losses for each timber-type combines the information from the AMBDEG summary and the stock tables. The dollar losses per cu. m. listed by diameter class for each species were imported to a spreadsheet containing the stock table for a particular timber type. In the stock table the volume per hectare in that timber type is broken down by species and diameter class. The volume figures in the stock table were multiplied by the dollar loss per cu. m. in the equivalent diameter class and species column. The result is a matrix showing the degrade loss in dollars. This matrix is then totaled giving a final figure for the degrade loss per hectare in that timber type; assuming 100% of that timber is actually attacked by *T. lineatum.* Table 5 shows the calculation of the total potential degrade loss for timber type 1 at a log degrade value of $500/cu.m. Similar calculations were performed for all other timber types in the study.

## Expanding the database (part 2)

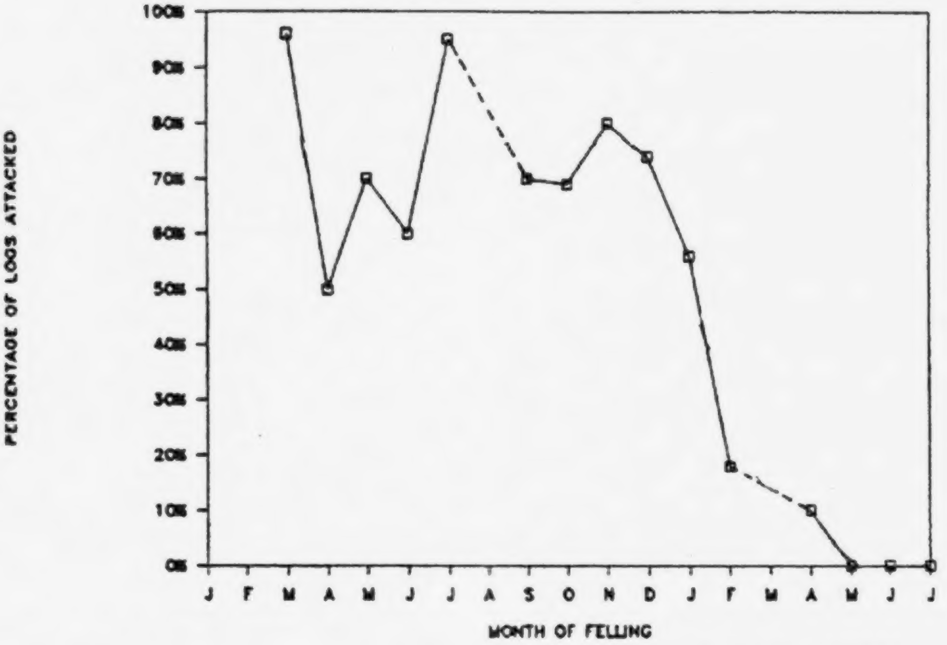
Table 6 shows the final expanded database structure with all the information used in the study. The total degrade field is based on a degrade value of $500/cu.m. The falling and logging dates represented in the database by the labels from the falling and logging themes were converted to a dd/mm/yy format. This was necessary to allow calculations based on these dates. A field was also added to display the current date, which can be updated monthly. The Fall_Date represents the date when the timber was felled, this is subtracted from the current date to estimate how long the timber has been down for. The difference between the two dates represents the Felled_Age which is listed in a separate field.

If logging in an area has not been completed then thetimber is considered to be at risk to attack by ambrosia beetles; in this case a "yes" is entered in the at risk column. The felled age is then calculated by subtracting the falling date from the current date. If logging in an area has been completed then the Felled_Age is calculated by subtracting the falling date from the logging date. This is useful if one wishes to design a Query to show areas that may have been attacked in previous years.

The Felled_Age, or length of time that the wood remains on the ground before being logged, is a very useful parameter. There are other serious age related effects on the quality of timber in addition to Ambrosia beetle attack, these include: weather checking, staining, and rotting (Mulock, 1986). The felled age parameter can be utilized in queries to show problem areas, and bring them to the attention of production personnel.

**Table 6: Expanded Database Structure**

| POLY AREA | POLY X | POLY Y | POLY STAT | GIS LINK | POLY PERIM | FC | LOGGING | FALLING | CURR DATE | FALL DATE | LOG DATE | LOG YEAR | FELLED AGE | SPECIES | AGE | HEIGHT | TYPE | VOLUME | ATTACK FA | TOT DGRADE | HAZARD | AT RISK | CALC LOSS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.41331 | 687119.80 | 5542481.27 | OK | 1 | 261.90 | SW50.23 | NL | FDEC90 | 02/02/91 | 12/15/90 | | 0 | 49 | FH | M | 5 | 21 | 1043 | 0.68 | 27991.48 | 19034.21 | Y | 7867.03 |
| 0.66825 | 687120.85 | 5542554.89 | OK | 2 | 490.86 | SW50.23 | NL | FNOV90 | 02/02/91 | 11/15/90 | | 0 | 79 | FH | M | 5 | 21 | 1043 | 0.75 | 27991.48 | 20993.61 | Y | 14028.98 |
| 1.15623 | 687209.72 | 5542489.22 | OK | 3 | 638.61 | SW50.21 | NL | FDEC90 | 02/02/91 | 12/15/90 | | 0 | 49 | FH | M | 5 | 21 | 1043 | 0.68 | 27991.48 | 19034.21 | Y | 22007.92 |
| 0.47383 | 687205.41 | 5542461.50 | OK | 4 | 509.79 | SW50.21 | NL | FNOV90 | 02/02/91 | 11/15/90 | | 0 | 79 | FH | M | 5 | 21 | 1043 | 0.75 | 27991.48 | 20993.61 | Y | 9943.21 |
| 0.08530 | 687134.39 | 5542392.70 | OK | 5 | 126.61 | SW50.21 | LNOV90 | FNOV90 | 02/02/91 | 11/15/90 | 11/15/90 | 1990 | 0 | FH | M | 5 | 21 | 1043 | 0.00 | 27991.48 | 0.00 | N | 0.00 |
| 0.00690 | 687130.17 | 5542380.25 | OK | 6 | 80.89 | SW50.21 | LOCT90 | FNOV90 | 02/02/91 | 11/15/90 | 10/15/90 | 1990 | 0 | FH | M | 5 | 21 | 1043 | 0.00 | 27991.48 | 0.00 | N | 0.00 |
| 0.05083 | 687131.96 | 5542370.49 | OK | 7 | 111.12 | SW50.21 | LOCT90 | FOCT90 | 02/02/91 | 10/15/90 | 10/15/90 | 1990 | 0 | FH | M | 5 | 21 | 1043 | 0.00 | 27991.48 | 0.00 | N | 0.00 |
| 0.30488 | 687174.90 | 5542388.48 | OK | 8 | 325.16 | SW50.2 | LOCT90 | FOCT90 | 02/02/91 | 10/15/90 | 10/15/90 | 1990 | 0 | HF | M | 4 | 20 | 368 | 0.00 | 7702.61 | 0.00 | N | 0.00 |
| 0.00217 | 687122.72 | 5542344.83 | OK | 9 | 29.54 | SW50.2 | LNOV90 | FOCT90 | 02/02/91 | 10/15/90 | 11/15/90 | 1990 | 31 | HF | M | 4 | 20 | 368 | 0.00 | 7702.61 | 0.00 | N | 0.00 |
| 0.14045 | 687156.00 | 5542355.63 | OK | 10 | 254.32 | SW50.2 | LNOV90 | FNOV90 | 02/02/91 | 11/15/90 | 11/15/90 | 1990 | 0 | HF | M | 4 | 20 | 368 | 0.00 | 2.61 | 0.00 | N | 0.00 |



(Daust, 1985)

| MONTH OF FELLING | ATTACK PROBABILITY (In Spring of Current Year) |
|---|---|
| Before December (Previous Year) | 0.75 |
| December (Previous Year) | 0.68 |
| January | 0.49 |
| February | 0.30 |
| March | 0.14 |
| April | 0.08 |
| May | 0.00 |

(Daust, 1985)

**Figure 3:** *Relationship between the probability of attack and Felling Month as assessed in the summer period.*

***The attack factor:*** Previous research has indicated that there is a very strong relationship between the date that the timber is felled and the level of ambrosia beetle attack. The percentage of logs attacked steadily decreases from trees felled in November (75% Attack), to those felled in April (0% attack) (Daust, 1985). Figure 3 shows the graphical relationship from which the Attack_Fac, or probability of attack was derived.

These values represent the levels of attack on logs that were assessed in the summer from June to August. For the purpose of this study anything felled in the year prior to the sampling period (ie. from January to December) was considered to have a probability of attack equal to .75. Logs felled from December of the previous year and the initial months of the current year (ie. sampling year) were assigned the probabilities listed in the table.

This Attack_Fac, or probability of attack was multiplied by the Total degrade values for those areas deemed to be at risk (ie. not yet logged). The resulting value is listed in the database field called Hazard. This figure represents the potential amount by which 1 hectare of a felled and bucked timber type may be degraded if it were not logged before April 1. This hazard value is multiplied by the individual polygon area for those polygons which are at risk, resulting in the actual calculated loss. The Calc_loss field can then be totaled for each individual setting. This determines the potential degrade that may be incurred if the wood is not removed before the Beetles fly.

### Query design

This inventory management system was set up to help identify those areas that are considered to be at risk to ambrosia beetle attack. In addition to identifying those areas an effort was made to classify them in terms of the value at risk, or the potential degrade loss. The areas were classified according to this hazard using a color coded system designed to attract the attention of pro-duction personnel.

In this application the following query criteria wereselected to identify those areas considered to be at risk to ambrosia beetle degrade:

- Poly_area between .1 ha. and 100 ha.
- At_Risk = "Y"
- Hazard between $0/ha. and $25,000/ha.

A solid shading pattern was chosen to highlight the polygons that met the query criteria. The colors chosen for the shading were as follows:

Green —> $0/ha. - $5000/ha. hazard
Purple —> $5000/ha. - $10,000/ha. hazard
Red —> $10,000/ha. - $15,000/ha. hazard
Orange —> $15,000/ha. - $20,000/ha. hazard
Yellow —> $20,000/ha. - $25,000/ha. hazard

## Results

Figure 4 and the associated table (Table 7) show the results for setting KT-126. The polygons shaded represent those areas that have been felled but not yet logged as of December 1990. The hazard rating shows the differences in the potential losses if these areas were indeed attacked by the beetles in the spring.

### Results summary

As of December 1990 setting KT-126 had a total of 45.89 hectares of timber at risk to ambrosia beetle attack. The potential loss was calculated to be $ 646,031.18

This type of information is useful in alerting the logging manager to the areas where problems associated with ambrosia beetle degrade loss may occur.

## Discussion

The results of the GIS query show large areas of timber considered to be at risk to ambrosia beetle attack. These figures are not necessarily representative of the actual degrade that may have occurred during the spring flight period. The data used in the study represented the progression of felling and yarding up to the beginning of December, 1990. Many of the areas identified as being at risk to ambrosia beetle attack may have been yarded prior to the spring flight period.

The intent of this paper was not to determine the extent of the losses suffered by CANFOR as a result of ambrosia beetles. The purpose of this research was to develop a system which could graphically display those areas considered to be at risk. These are the areas which may indeed suffer degrade if actions are not taken to remove the wood prior to the spring flight period of *T. lineatum.*

Logs will be attacked at a maximum time of 60 days from felling. Where epidemic populations exist attack may be even sooner. Some form of incentive to protect the quality and value of the timber being logged is necessary. The extra costs involved in better planning and inventory management can be easily offset by maintaining log quality throughout all phases of the harvesting process.

## Conclusions

Management of setting inventory using GIS gives production personnel a powerful tool which can aid scheduling in attempts to reduce ambrosia beetle attack. The progress of felling and logging can be instantly accessed and tracked on a setting by setting basis. The value of the timber can be instantly assessed and an estimate of the potential degrade can be acquired, thus giving managers the information needed to make cost saving decisions.

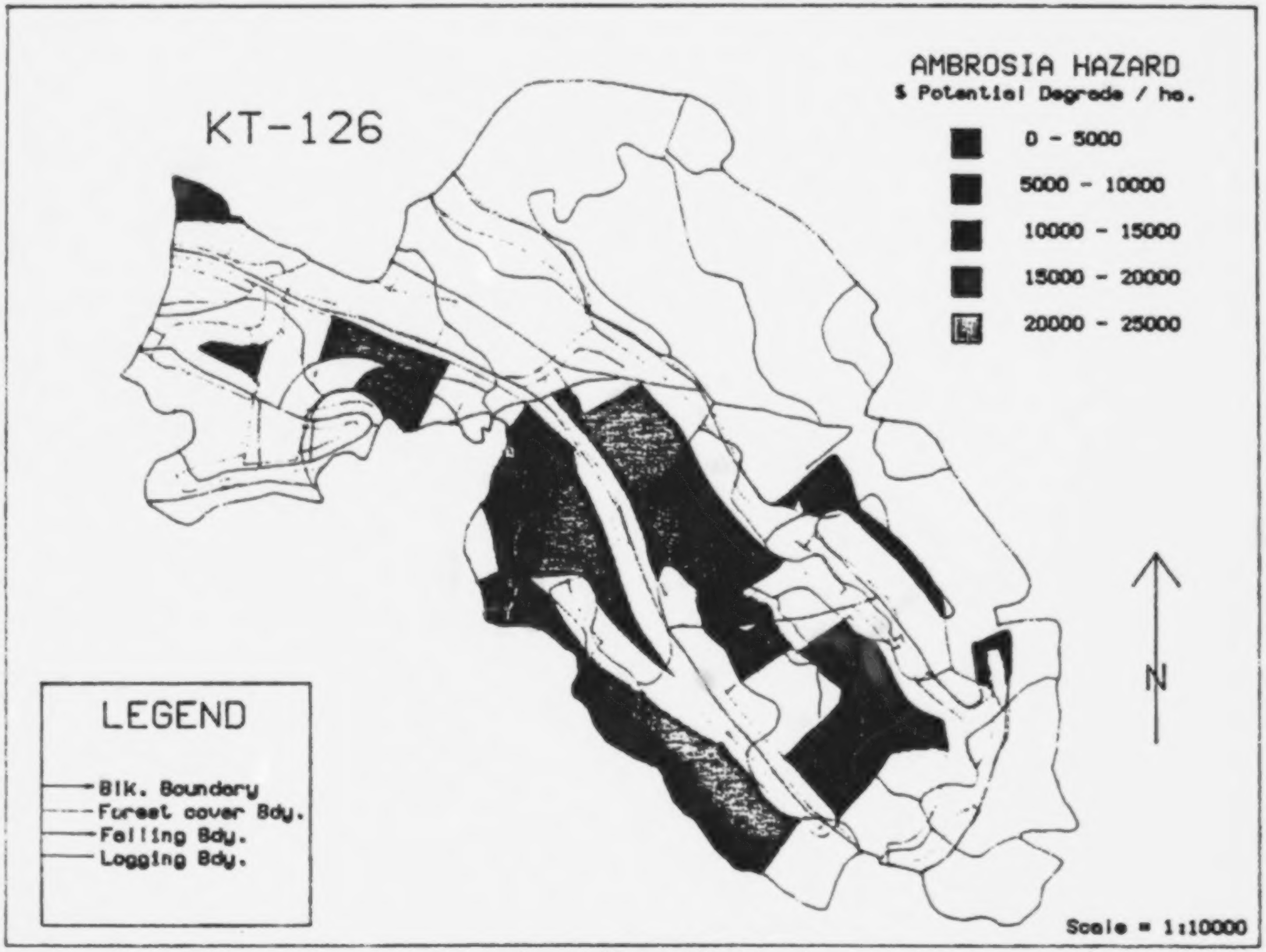

KT-126
AMBROSIA HAZARD
$ Potential Degrade / ha.
0 - 5000
5000 - 10000
10000 - 15000
15000 - 20000
20000 - 25000
N
LEGEND
Blk. Boundary
Forest cover Bdy.
Falling Bdy.
Logging Bdy.
Scale = 1:10000

Figure 4: *KT - 126, Ambrosia hazard*

Potential Degrade Summary

| GIS_LINK | POLY_AREA ha. | TYPE CODE | VOLUME Cu.m./ha | FOREST COVER | LOGGING | FALLING | CURR_DATE | FALL_DATE | FELLED AGE | AT_RISK | TOT_DGRADE $/ha. | ATTACK_FAC | HAZARD $/ha. | CALC_LOSS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 79 | 0.99146 | 99 | 903 | KT126.1 | NL | FSEPT89 | 02-Feb-91 | 15-Sep-89 | 505 | Y | $23,099.02 | 0.75 | $17,324.26 | $17,176.31 |
| 102 | 0.87042 | 99 | 903 | KT126.2 | NL | FMAY89 | 02-Feb-91 | 15-May-89 | 628 | Y | $23,099.02 | 0.75 | $17,324.26 | $15,079.39 |
| 136 | 0.59521 | 99 | 903 | KT126.1 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $23,099.02 | 0.75 | $17,324.26 | $10,311.58 |
| 206 | 0.80067 | 99 | 903 | KT126.2 | NL | FJUNE90 | 02-Feb-91 | 15-Jun-90 | 232 | Y | $23,099.02 | 0.75 | $17,324.26 | $13,871.02 |
| 210 | 0.31046 | 83 | 860 | KT126.3 | NL | FJULY90 | 02-Feb-91 | 15-Jul-90 | 202 | Y | $20,821.26 | 0.75 | $15,615.94 | $4,848.13 |
| 213 | 1.55119 | 99 | 903 | KT126.2 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $23,099.02 | 0.75 | $17,324.26 | $26,873.22 |
| 229 | 1.27753 | 83 | 860 | KT126.3 | NL | FJUNE90 | 02-Feb-91 | 15-Jun-90 | 232 | Y | $20,821.26 | 0.75 | $15,615.94 | $19,949.84 |
| 245 | 0.18484 | 83 | 860 | KT126.3 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $20,821.26 | 0.75 | $15,615.94 | $2,886.45 |
| 317 | 0.60909 | 67 | 477 | KT126.4 | NL | FNOV90 | 02-Feb-91 | 15-Nov-90 | 79 | Y | $10,902.63 | 0.75 | $8,176.98 | $4,980.51 |
| 319 | 2.11292 | 67 | 477 | KT126.4 | NL | FFEB90 | 02-Feb-91 | 15-Feb-90 | 352 | Y | $10,902.63 | 0.31 | $3,379.82 | $7,141.28 |
| 329 | 0.79866 | 99 | 903 | KT126.5 | NL | FFEB90 | 02-Feb-91 | 15-Feb-90 | 352 | Y | $23,099.02 | 0.31 | $7,160.70 | $5,718.96 |
| 332 | 0.12790 | 83 | 860 | KT126.6 | NL | FNOV90 | 02-Feb-91 | 15-Nov-90 | 79 | Y | $20,821.26 | 0.75 | $15,615.94 | $1,997.28 |
| 333 | 4.45026 | 99 | 903 | KT126.5 | NL | FSEPT89 | 02-Feb-91 | 15-Sep-89 | 505 | Y | $23,099.02 | 0.75 | $17,324.26 | $77,097.47 |
| 334 | 2.57063 | 83 | 860 | KT126.6 | NL | FFEB90 | 02-Feb-91 | 15-Feb-90 | 352 | Y | $20,821.26 | 0.31 | $6,454.59 | $16,592.36 |
| 340 | 0.43959 | 67 | 477 | KT126.4 | NL | FSEPT89 | 02-Feb-91 | 15-Sep-89 | 505 | Y | $10,902.63 | 0.75 | $8,176.98 | $3,594.52 |
| 341 | 0.20975 | 83 | 860 | KT126.6 | NL | FSEPT89 | 02-Feb-91 | 15-Sep-89 | 505 | Y | $20,821.26 | 0.75 | $15,615.94 | $3,275.45 |
| 345 | 0.14618 | 99 | 903 | KT126.5 | NL | FMAY89 | 02-Feb-91 | 15-May-89 | 628 | Y | $23,099.02 | 0.75 | $17,324.26 | $2,532.46 |
| 348 | 5.56954 | 99 | 903 | KT126.5 | NL | FJULY89 | 02-Feb-91 | 15-Jul-89 | 567 | Y | $23,099.02 | 0.75 | $17,324.26 | $96,488.17 |
| 362 | 0.49427 | 83 | 860 | KT126.6 | NL | FSEPT89 | 02-Feb-91 | 15-Sep-89 | 505 | Y | $20,821.26 | 0.75 | $15,615.94 | $7,718.49 |
| 380 | 3.41639 | 99 | 903 | KT126.7 | NL | FSEPT89 | 02-Feb-91 | 15-Sep-89 | 505 | Y | $23,099.02 | 0.75 | $17,324.26 | $59,186.44 |
| 381 | 1.84473 | 99 | 903 | KT126.7 | NL | FFEB90 | 02-Feb-91 | 15-Feb-90 | 352 | Y | $23,099.02 | 0.31 | $7,160.70 | $13,209.55 |
| 384 | 0.72793 | 99 | 903 | KT126.7 | NL | FJULY89 | 02-Feb-91 | 15-Jul-89 | 567 | Y | $23,099.02 | 0.75 | $17,324.26 | $12,611.09 |
| 396 | 0.73195 | 99 | 903 | KT126.7 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $23,099.02 | 0.75 | $17,324.26 | $12,680.50 |
| 399 | 2.93914 | 77 | 802 | KT126.14 | NL | FJULY89 | 02-Feb-91 | 15-Jul-89 | 567 | Y | $19,598.79 | 0.75 | $14,699.09 | $43,202.60 |
| 400 | 0.88592 | 77 | 802 | KT126.8 | NL | FJULY89 | 02-Feb-91 | 15-Jul-89 | 567 | Y | $19,598.79 | 0.75 | $14,699.09 | $13,022.22 |
| 403 | 1.09703 | 77 | 802 | KT126.8 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $19,598.79 | 0.75 | $14,699.09 | $16,125.34 |
| 413 | 0.31895 | 77 | 802 | KT126.13 | NL | FJULY89 | 02-Feb-91 | 15-Jul-89 | 567 | Y | $19,598.79 | 0.75 | $14,699.09 | $4,688.28 |
| 426 | 0.13121 | 77 | 802 | KT126.8 | NL | FSEPT89 | 02-Feb-91 | 15-Sep-89 | 505 | Y | $19,598.79 | 0.75 | $14,699.09 | $1,928.67 |
| 433 | 0.10431 | 77 | 802 | KT126.13 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $19,598.79 | 0.75 | $14,699.09 | $1,533.26 |
| 435 | 0.12674 | 83 | 860 | KT126.16 | NL | FFEB90 | 02-Feb-91 | 15-Feb-90 | 352 | Y | $20,821.26 | 0.31 | $6,454.59 | $818.06 |
| 438 | 1.28037 | 77 | 802 | KT126.8 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $19,598.79 | 0.75 | $14,699.09 | $18,820.27 |
| 442 | 0.11567 | 77 | 802 | KT126.13 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $19,598.79 | 0.75 | $14,699.09 | $1,700.25 |
| 443 | 0.73742 | 75 | 667 | KT126.19 | NL | FJULY89 | 02-Feb-91 | 15-Jul-89 | 567 | Y | $17,067.54 | 0.75 | $12,800.66 | $9,439.46 |
| 444 | 0.39091 | 75 | 667 | KT126.19 | NL | FFEB90 | 02-Feb-91 | 15-Feb-90 | 352 | Y | $17,067.54 | 0.31 | $5,290.94 | $2,068.28 |
| 448 | 3.42540 | 77 | 802 | KT126.12 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $19,598.79 | 0.75 | $14,699.09 | $50,350.26 |
| 454 | 1.82442 | 79 | 754 | KT126.9 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $17,960.33 | 0.75 | $13,470.25 | $24,575.39 |
| 470 | 0.33267 | 79 | 754 | KT126.15 | NL | FJULY89 | 02-Feb-91 | 15-Jul-89 | 567 | Y | $17,960.33 | 0.75 | $13,470.25 | $4,481.15 |
| 488 | 0.35024 | 75 | 667 | KT126.19 | NL | FJULY89 | 02-Feb-91 | 15-Jul-89 | 567 | Y | $17,067.54 | 0.75 | $12,800.66 | $4,483.30 |
| 491 | 0.18011 | 79 | 754 | KT126.15 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $17,960.33 | 0.75 | $13,470.25 | $2,426.13 |
| 493 | 0.15781 | 79 | 754 | KT126.15 | NL | FJULY89 | 02-Feb-91 | 15-Jul-89 | 567 | Y | $17,960.33 | 0.75 | $13,470.25 | $2,125.74 |
| 516 | 0.65787 | 75 | 667 | KT126.19 | NL | FAUG89 | 02-Feb-91 | 15-Aug-89 | 536 | Y | $17,067.54 | 0.75 | $12,800.66 | $8,421.17 |
| Total = | 45.88782 | | | | | | | | | | | | Total = | $646,031.18 |

Table 7: Setting KT-126

CANFOR currently keeps track of the Felling and Logging in each active setting. The areas are updated monthly and traced onto the maps retained in the active setting files. At the end of each month the linework is digitized and the areas are recorded in the month end reports.

Switching from the present system, where Felled and Bucked areas are digitized for month end reports, to a system supported by GIS would require little additional work. The linework could be digitized directly into Terrasoft on a setting by setting basis at the end of each month. This way the linework would be retained and processed. Areas could still be obtained and recorded in the month end reports.

With the database constructed the system could easily be updated monthly.This would only require processing of the new linework, updating the database, changing the current date, and re-performing the calculations. The database could then be queried on demand using the same query design. The result would be a system that tracks Felled and Bucked inventory showing areas of priority logging.

Cleaning up logging activities will benefit any timber harvesting operation. It will result in more high quality, and valuable logs reaching the mill.

## Literature cited

Anon., 1990. Managing Ambrosia Beetles With Semiochemicals. *Phero Tech Inc. Technical Bulletin* Phero Tech Inc. pp 4

Daust, Kurt D. 1985. The Influence of Site Characteristicsand Logging Techniques on Ambrosia Beetle Damage in the Lake Cowichan Area. *Bachelor of Science in Forestry Thesis.* April 7, 1985. pp 68.

Gray, D.R. and J.H. Borden, 1985. Ambrosia beetle attack on logs before and after processing through a dryland sorting area. *Forestry Chronicle.* 61: pp 299-302.

Johnson, Norman E. 1958. Ambrosia Beetle Infestation of Coniferous Logs on Clearcuttings In Northwestern Oregon. *Weyerhaeuser Forest Research Note # 58.* pp 508-511.

McLean, John A. 1991. Personal Correspondance.

McLean, John A. 1990. AMBDEG: A Program to Estimate Value Losses in Ambrosia Beetle Infested Conifer Logs. Version 3.0b September 16, 1990.

McLean, John A. 1988. Ambrosia Beetles in the Forests of the Northwest. *Northwest Environmental Journal.* 4:2. pp 324-327.

McLean, John A. and Stokkink, E. 1988. Challenges in Implementing Ambrosia Beetle Pest Management Programs in British Columbia. *XVII International Congress of Entomology Symposium*, "Integrated Control of Scolytid Bark Beetles". pp 179-187.

Mulock, Ed. 1986. Log Value Losses To F&B, Windfall and Fringe Damaged Timber. *Canadian Forest Products Inc. Internal Report.* August 29, 1986. pp 2.

Nijholt, W.W. 1978. Ambrosia Beetle: A Menace to the Forest Industry. *Canadian Forest Service, Pacific Forest Research Center Report.* BC-P-25. pp 8.

Shore et al., 1987. Reproduction and Survival of the Ambrosia Beetle *Trypodendron lineatum* (Oliv.) (Coleoptera: Scolytidae) In Douglas Fir and Western Hemlock Logs. *Canadian Entomologist.* 119. pp 131-139

Wilson, E.A. 1988. Ambrosia Beetles the Opportunity for Added Profit. *Canadian Forest Products Inc. Internal Report.* February, 1988. pp 9.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# Recent developments in spatial analysis of fisheries acoustic data

**R. Kieser**
**K. Cooke**
Pacific Biological Station,
Hammond Bay Road,
Nanaimo, B.C., V9R 5K6.

**G. Langford**
Geo-Spatial Systems Ltd.
102 Sir Lancelot Place,
Nanaimo, B.C., V9T 4K4

## Abstract

*Fisheries resource managers and researchers use acoustic methods, such as sonar and echosounding, to examine the distribution and behaviour patterns of various fish stocks, identify suitable fish habitat, and estimate stock abundance. In the past it has been difficult to capture the geographic nature of fisheries data such as catch information and acoustic fish density measurements and to visualize and quantify their relation with oceanographic and environmental data. Considerable progress has been made in this area by using the COMPUGRID Spatial Analysis System to display, analyze and quantify fisheries acoustic and related data. A case study is presented to illustrate the new analysis procedures.*

## Introduction

Acoustic surveys generally proceed along transects and record qualitative and quantitative data in the form of echograms and fish density measurements. The echogram provides a visual presentation of fish concentrations and ocean bottom (Figure 1). Fish density can be obtained by measuring the returned echo power (Mitson 1983). Recently Kieser and Langford (1991) have reported a GIS based analysis of the data shown in the echogram that recognizes and quantifies individual fish schools. In this paper we describe the use of GIS tools to analyze transect based acoustic survey data.

This approach allows us, for the first time, to fully visualize the spatial characteristics of acoustic data, bathymetry, oceanographic information and biological sampling data. These data sets are stored as multiple files that are treated as independent and complementary layers. Maps of fish density and habitat can now be created. This approach also provides the framework for detailed quantitative analyses relating fish distributions to their environment.

By incorporating our data into the COMPUGRID Spatial Analysis System (Langford 1991), we obtain geographically referenced displays of transects and surface density distributions. Major and minor study areas can easily be defined and observed fish concentrations can be extrapolated to the entire area of interest. A sample data set from a hake (*Merluccius productus*) stock assessment cruise, off Vancouver Island, August, 1991 (Saunders et. al. 1992) is used to illustrate this new GIS application.

## Data collection

The data were collected with a BioSonics scientific echo sounder system consisting of a transducer, transceiver, chart recorder, echo integrator, and an analog tape recorder (Kieser 1985; Kieser et. al. 1987; Mitson 1983; Saunders et. al. 1992). The echo sounder transmits a 38 kHz, 0.6 ms, 1000 W pulse with a repetition rate of 1 Hz. Its receiver employs a time varied gain to compensate for spreading and absorption losses. The system generates an echogram (Figure 1) which provides an effective visual presentation of fish distributions and ocean bottom. Echo integration reduces the data presented in Figure 1 to surface density measurements that are calculated for each one minute interval along the transect. A PC type computer is used to store the output from the echo integrator and to process all data.

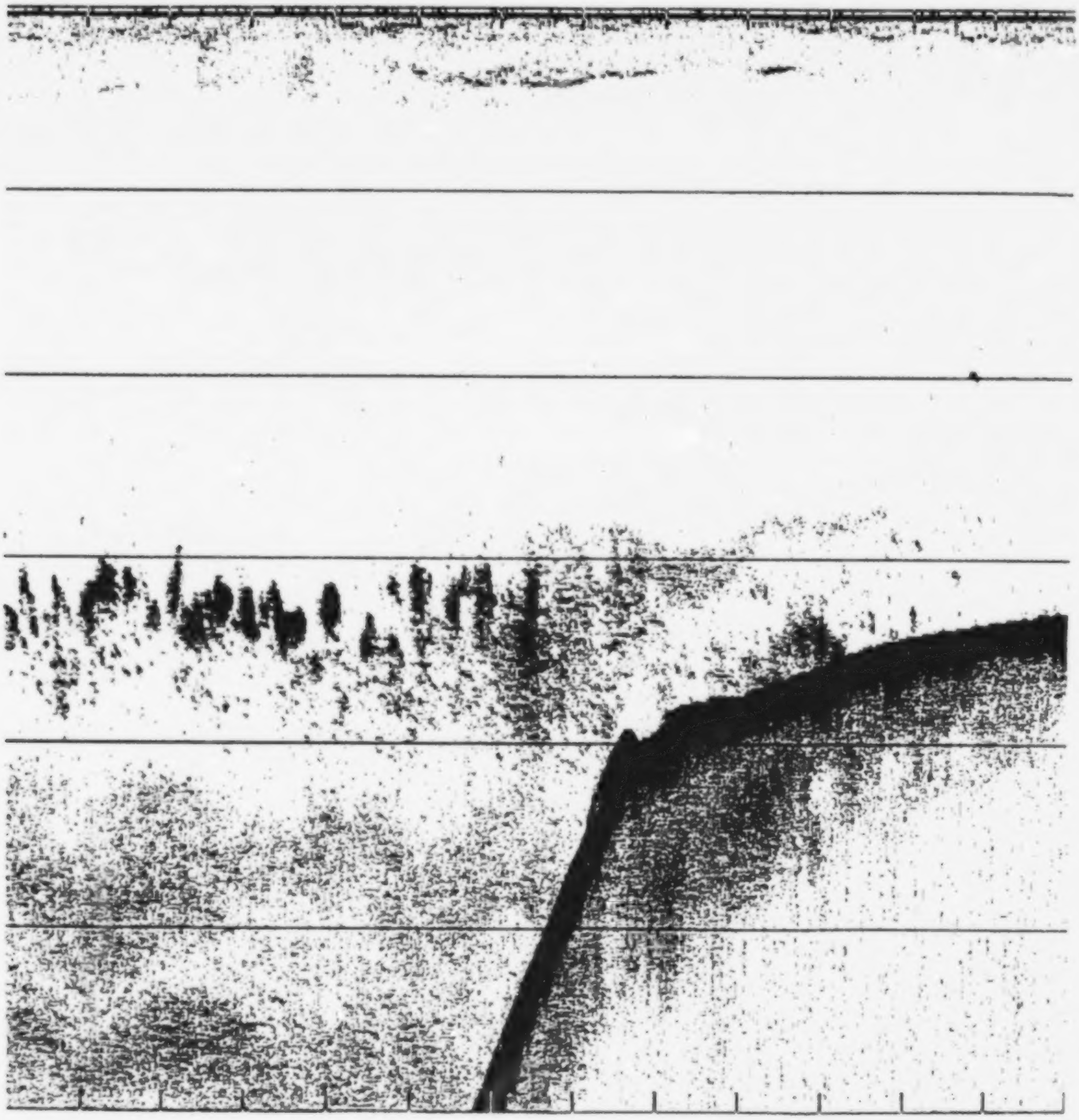


***Figure 1:*** *Echogram with typical hake schools at the edge of the continental shelf off Vancouver Island. The depth range is 0 to 300 m, the 10 minute recording covers a distance of 2.6 km.*

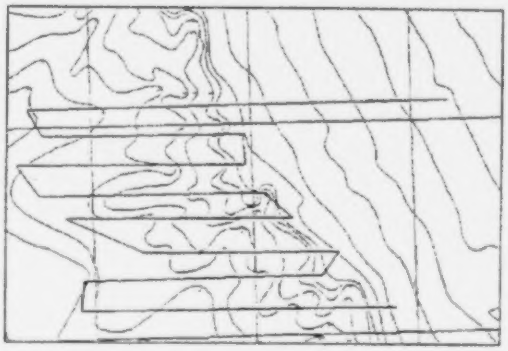


*Figure 2. Transect plot with latitude, longitude grid, and bathymetry. Grid spacing is 0.1°. Depth increases toward the West, 10 m isobaths from 110 to 200 m and 100 m isobaths from 200 to 800 m are shown. OVERLAY has been used to combine the data sets.*

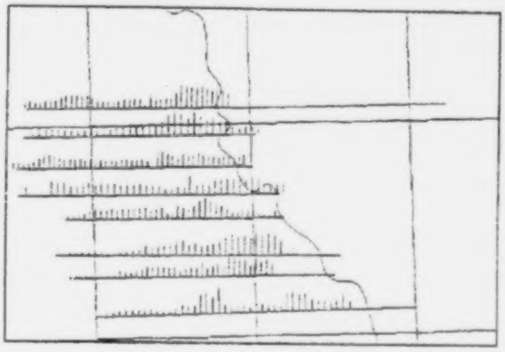


*Figure 3. Surface density bar plot. The height of the bars represents surface densities from 0.001 to 0.1 kg/m$^2$. Latitude, longitude grid spacing is 0.1°. The 180 m isobath is shown to indicate the shelf break.*

## Data analysis procedure

### Conventional analysis of hydroacoustic data

Conventional analysis procedures result in a series of primitive transect and surface density plots and a biomass estimate. These products lack accurate geographic referencing and individual plots cannot be related to each other or to other important information. To provide additional display and analysis capabilities, the acoustic software was modified to output latitude, longitude and surface density in ASCII files for input to the GIS. Conventional processing of the hydroacoustic data includes geographic referencing and editing. Latitude and longitude coordinates are assigned to each surface density measurement. Data editing is required to choose density measurements from selected fish schools and to exclude noise, ocean bottom and echoes from unwanted sources in the water column.

### GIS data input

Custom software is used to convert the latitude, longitude, surface density ASCII files to the format that is used by the COMPUGRID Spatial Analysis System. The process allows us to choose an appropriate geographic area and scales the surface densityvalues.

Several ancillary data sets are required to produce a composite picture of the fish distributions of interest. These include, bathymetry, trawl haul locations and oceanographic parameters. Bathymetry and other cartographic information are obtained by digitizing printed charts or from digital charts that are available from the Canadian Hydrographic Service or commercial sources. Simple custom procedures have been developed to input trawl haul locations and oceanographic data.

### Transect and surface density bar plot

The first result from our GIS based processing is a map of the transects that were executed (Figure 2). Transects are shown with the appropriate latitude, longitude grid and bathymetry. Bathymetry is important because fish schools are often found in association with features such as slopes, canyons or the continental shelf break. Figure 2 illustrates the overlaying of transects and bathymetry, other data can easily be included.

The surface density bar plot (Figure 3) shows one of several possible options to visualize the fish distribution. Each upright in Figure 3 is placed at the appropriate location along the transect, with its height representing the surface density. A density range from 0.001 to 0.1 kg/m$^2$ is displayed on a logarithmic scale. Alternatively, the measured surface density values can be presented by dot size or coloured dots. The latter presentation can be quite effective when a blue-green-red colour scale is used with relatively few classes.

### Surface density matrix

The surface density COMPUGRID file is the basis for quantitative processing that yields estimates of the survey

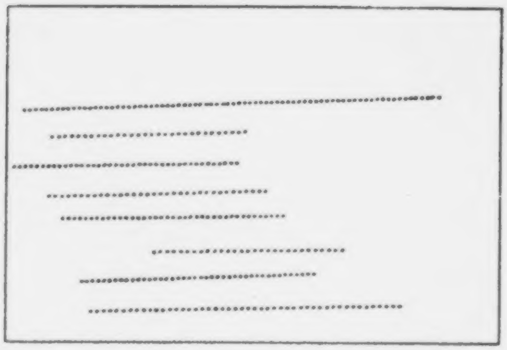


*Figure 4. The surface density matrix is the basis for quantitative analysis. Each measurement occupies a single matrix element $z_{ij}$ at the location of the surface density measurement. Dots have been enlarged for better visibility.*

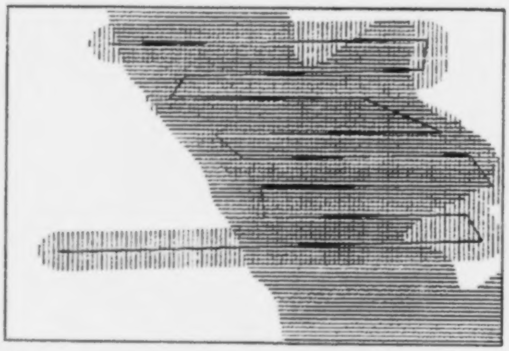


*Figure 5. Composite mask to define the active survey area. Vertical bars indicate the area buffered from the surface density matrix, horizontal bars are limited by the 160 and 600 m isobath. Only matrix elements that satisfy both conditions belong to the survey area.*

area and fish biomass. It is a two dimensional matrix with appropriately placed surface density values $z_{ij}$ (Figure 4). Additional information such as transect labels are excluded to avoid interference during quantitative processing.

At nine knots survey speed the distance between surface density measurements made at one minute intervals is 278 m. Given a maximum matrix dimension of 2048 by 2048 for the COMPUGRID file this yields a maximum survey area of 569 by 569 km. This is much larger then typical survey area dimensions of 100 by 50 km. COMPUGRID values for $z_{ij}$ have a range from 0 to 32000, while surface density values range from $10^{-4}$ to 1 kg/m$^2$. Generally we multiply the densities by $10^4$, this allows us to present their entire range as integers from 1 to $10^4$.

### Area expansion and biomass estimation

We use the term area expansion to describe the process of expanding the surface density measurements to the entire survey area. In our conventional hydroacoustic software this is accomplished by expanding the values that were measured along the transect to fill the space between transects. Each surface density covers an area that is bounded on either side by lines midway between the present and adjacent transects and by equidistant lines between the present and adjacent surface density locations. For a simple rectangular transect pattern this approach is equivalent to PROXIMAL analysis (Pavlidis 1982, Tomlin 1990). Each unknown point will assume the value of the nearest measured point. Proximal analysis has the advantage that it can be used with irregular transects.

Alternatively, several interpolation algorithms are available which can produce pleasing images of the biomass distribution. It is important to realize that the final biomass estimate will depend on the particular interpolation algorithm that is used. Because of its simplicity, we prefer the proximal algorithm. Future work is designed to look for additional information that will allow us to define a more complex area expansion algorithm and its parameters.

Area expansion begins by constructing a mask to delineate the survey area. Three steps are required to generate this mask. First BUFFER is used to expand measured surface density locations to cover the area between transects. Next the bathymetry data are overlaid to eliminate areas of inappropriate depth (Figure 5). As a last step the drawing tools provided by the GIS can be used to fill in missed areas or crop unwanted portions.

Proximal analysis is used to expand the measured surface densities (Figure 4) to the survey area defined by the mask (Figure 5). Figure 6 shows a black and white presentation that employs four surface density classes to give an impression of the biomass distribution. A biomass estimate for the survey area is obtained by summing all matrix elements and multiplying by the area per element and a scaling factor to compensate for the integer nature of the $z_{ij}$ values.

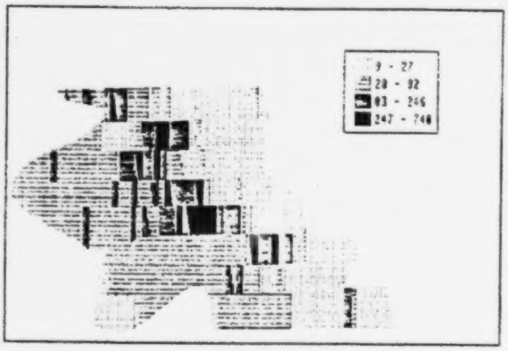


***Figure 6.*** *PROXIMAL has been used to generate surface density values for all matrix elements in the active survey area defined in Figure 5. The resulting COMPUGRID file has been classified into four logarithmic classes to generate a visual presentation. Some boundaries between density classes are sloped, this is a consequence of the finate resolution of raster based processing. The effect on the results is insignificant as long as several empty matrix elements are between each pair of the surface density values.*

## Discussion

The paper describes a new approach for the display and analysis of fisheries acoustic data that uses standard GIS tools. This approach brings a range of new capabilities and analysis procedures to the interpretation of fisheries acoustic and related data. Some of these procedures such as PROXIMAL and BUFFER were virtually unknown in this field. Others such as OVERLAY of multiple data sets were too cumbersome to be practical.

We have chosen a raster based GIS to satisfy our quantitative requirements that demand the analysis not only of irregularly spaced fish density measurements but also of multi dimensional data sets that are sampled on different grids. We find that the accuracy of locating a measurement within the data matrix is sufficient even when large areas are analyzed. At the same time, our raster based GIS provides an excellent compromise between analytical and display capabilities. The typical graininess of the display does not effects the quality or appeal of our results.

Although a menu driven user interface is available, we generally use batch processing for our routine work. This results in a substantial increase in throughput and provides comprehensive documentation of the analysis procedures and parameters that were used. Documentation of procedures is essential when biomass estimates are used to determine the total allowable catch of a particular fish species.

The analysis described above depends on the use of standard GIS tools. Several research applications have emerged from this work that explore the relationship between fish aggregations and their environment, or that address the automatic detection and analysis of fish schools (Kieser and Langford, 1991). This work would not have been possible without assimilating the ideas and software tools that are part of the GIS approach.

## References

Kieser, R. (1985). A versatile echointergration/data logging system for fisheries research and management. Can. Tech. Rep. Fish. Aquat. Sci. 1378:48.

Kieser, R., and G.J. Langford, (1991). An application of spatial analysis to fisheries acoustics. In: Proceedings of GIS'91 Symposium, Vancouver, B.C., pp. 335-339.

Kieser, R., T.J. Mulligan, M.J. Williamson, and M.O. Nelson, (1987). Intercalibration of two echo integration systems based on acoustic backscattering measurements. Can. J. Fish. Aquat. Sci. 44:562-572.

Langford, G.J. (1991). COMPUGRID User Manual. Geo-Spatial Systems Ltd., Nanaimo, B.C., 186 p.

Mitson, R.B. (1983). Fisheries Sonar. Fishing News Books Ltd., Farnham, Surrey, England, 287 p.

Pavlidis, T. (1982). Algorithms for Graphics and Image Processing. Computer Science Press, Rockville, M.D.

Saunders, M., R. Kieser, and K.D. Cooke, (1992). W.E. Ricker Hake (*Merluccius productus*) hydroacoustic stock assessment cruise off the westcoast of Vancouver Island, August, 1991. Can. Man. Rep. of Fish. Aquat. Sci. In press.

Tomlin, C.D. (1990). Geographic Information Systems and Cartographic Analysis. Prentice Hall, Englewood Cliffs, N.J.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# The next generation of digital spatial technologies: error utilization, thematic map and remote sensing integration, spatial operators, and data structures

**Kim Lowell**
**Christopher Gold**
**Geoffrey Edwards**

Industrial Research Chair in Geomatics
Center for Research in Geomatics
Universite Laval, Pavillon Casault
Cite Universitaire, Ste-Foy G1K 7P4
Quebec, Canada
Tel: (418) 656-2530
Fax: (418) 656-7411

## Abstract

*It is suggested that new ways of thinking about spatial analysis are required to optimize the use of digital technologies. "Real-world-to-map" simplifications must be quantified and incorporated into spatial analysis with particular emphasis on boundary and category error. This could also lead to a better way to integrate thematic map-based and remotely-sensed data. However, such developments will require the creation of a new family of spatial operators which work on "error" or "uncertainty" surfaces rather than categorical thematic maps. These operators should work across maps and polygon boundaries rather than by considering absolute location to be the primary characteristic of spatial entities. The use of these concepts in digital spatial technologies may also require the development of an alternative data structure. The Voronoi diagram is presented and discussed as a possible solution.*

## Introduction

Thematic maps show the categories for a particular "theme" over a given area. For example, a thematic map might indicate the population density (persons per unit area) within each "township" for a region of interest. The density could either be shown in "raw values" — 2.2 people per hectare — or categorically — 1 to 3 people per hectare with the township polygons color-coded accordingly. Either way, such a map would show definite values for distinct areas, and polygons would be assumed to be uniform within. Finally, boundaries between townships would be mapped as (dimensionless) lines meaning that changes in population density between townships would occur instantaneously.

That such geographical representations have existed for a long time is not surprising. Indeed, most people are familiar with such storage formats for spatial data. However, it is this familiarity which makes one forget some of the inherent limitations in such maps. One is simply the constraints which are placed on any cartographer by "pen and paper technology." That is, using lines as delimiters of "meaningful" polygons is an easy way to represent "the world" on a piece of paper. However, this conventional thematic map representation has also evolved to its present form due to data limitations. In the preceding example, one might actually desire population density by "section of townships." However, if population records have not been kept for every "section," one cannot produce such a map. Thus one must group units spatially until one has a datum for each polygon — townships in this case. This also makes it evident that data availability will determine the size of the minimum mapping unit. Finally, existing thematic map representations are also employed partly because humans have a limited ability to comprehend complex spatial surfaces. Any person who has taught the interpretation of contour maps to first-year students knows the difficulties inherent in enabling students to understand even that familiar, and relatively simple, geographical representation.

Relative to the use of GIS and remote sensing technologies (which together we call Digital Spatial Technologies or DSTs) for natural forests, there are numerous implications of traditional map representations and their inherent limitations. In forest management, a considerable amount of effort is expended to produce "cover type" maps, "site quality" maps, etc. In creating such maps, polygon boundaries are treated as being dimensionless, identifiable, and locatable on the ground. Furthermore, polygons are treated as if they are completely uniform within (though one might attach confidence limits to the attributes of a given polygon). Overlays of the polygon boundaries of separate feature-maps are done deterministically using Boolean logic meaning that queries of resulting maps will indicate that a point is either definitely "in" or "out" of a given polygon. Moreover, confidence limits for totals (for volume over an area, for example) are based on statistical sampling errors exclusively, and not on potential errors in boundary location and/or category identification.

It is apparent that in certain endeavors, the topics mentioned in the previous paragraph are not an issue. In human geography, for example, boundaries may be, in fact, dimensionless, identifiable, and locatable. However, making the assumption that, in the mapping of natural resources, boundaries can be represented the same way may lead to a number of gross oversimplifications. Firstly, in any natural system, polygon boundaries are not definite, nor are they unchanging over time. For example, the difference between a Birch(Spruce) species mix and a Spruce(Birch) mix is relatively subtle. Thus the identification and location of a boundary between similar types has a high potential for error. Furthermore, because one of these types may gradually evolve into the other, the boundary is likely to change over time. Secondly, polygons of natural features are not uniform within. Thus to treat an area in the center of a mature forest polygon, for example, as definitely having the same volume as an area closer to the edge of a scrubland may be a questionable procedure.

We believe that a number of issues must be addressed relative to the use of DSTs in the management of natural forests. The purpose of this paper is to present these issues for consideration, discuss them in some detail, and propose possible solutions. Specifically, we will focus on four subjects: new ways to measure and use spatial errors, an alternative map representation to aid thematic map and remote-sensing integration, the creation of "probabilistic" instead of deterministic spatial operators, and a suitable data structure for the incorporation of these ideas into digital spatial analysis.

## Measurement and use of spatial errors

Selected spatial errors have been the subject of considerable DST research. These have focussed largely on digitizing errors in GIS and/or geometric correction and geographic registration in remote sensing. However, the study of digitizing errors reflects largely an implicit assumption that features being digitized are identifiable and locatable. Thus the study of such errors — including those of geometric correction in remote sensing — "make sense" in subjects such as human geography and plantation forestry. But in natural forest stands, the very existence of a boundary line may be in doubt, not merely its location. Considering these factors, digitizing and registration errors are likely to comprise a relatively small amount of the "total error" in the conversion of natural spatial phenomena to a digital map product (Chrisman 1982).

The authors of this paper have identified two types of errors that we believe are more important in natural forest stands than digitizing and registration errors. We refer to these as "boundary" and "category" errors. Rather than being thought of as "mistakes" these should be interpreted simply as real-world-to-map "simplifications." Generally, one assumes that such errors are merely inherent in map-based data, and, therefore, ignorable. Or that, because they are difficult to quantify and use, they must be assumed to be non-existent. The implications of such assumptions can be severe.

For example, suppose that in a forest there is a swampy area directly beside a highly productive forest type. On a map of the area, the boundary between the two would be drawn as a line thus suggesting that each type can actually be precisely delimited. In "the real world," however, the boundary between the two would actually be a broad transition zone. If the cartographer consistently estimated that there is less swamp than actually exists by placing the boundary in the wrong place, forest planning efforts could be unreliable. Specifically, one would overestimate the amount of volume on the forest because the cartographer had overestimated the amount of high-productivity land.

This is one issue that we are trying to address in DST research. It can be understood intuitively that some natural boundaries will be relatively easy to identify and locate. One example is the boundary between a clearcut and a mature forest. Similarly, some individual categories of stands are easy to identify such as pure, even-aged stands of jack pine *Pinus banksiana Lamb* in Quebec. However, even with pure forest stands, boundaries become difficult to identify if these stands are surrounded by a highly variable natural type such as "scrubland." Recognition of this factor suggests that there may be a relationship between boundary and category errors.

Two projects are underway to attempt to identify and quantify both category and boundary errors and to establish a relationship between the two. In one project, within a boreal forest in Quebec 64 sample plots have been established on an 8-by-8 grid with plots spaced 50 metres apart. The dbh of all trees which are within 5.64 m of plot center and which are 10 cm dbh and greater have been measured. Similarly, the dbh of all trees within 3.57 m of plot center which are 1 to 10 cm dbh have been measured. Finally, the heights and dbh's of three dominant or co-dominant trees have been measured at each plot. This grid of sample plots is in the process of being located on a photo-interpreted map of forest cover types. From this, we intend to determine if there is a relationship between humanly-interpreted aerial photographs and the ground-based forest inventory information. Because of the relatively tight spacing of plots, we should be able to identify and quantify points at which both boundary and category errors occur. These data will also be used with remotely sensed digital data collected from a sensor mounted on a fixed-wing aircraft. With this information we intend to evaluate the ability of high-resolution spectral data to detect forest type boundaries relative to the characteristics of each forest type.

The second project involves synthetic remote-sensing data generated from a computerized image generator. We believe the magnitude of boundary and category errors will be found to be related to the difference in the means and variances between two adjacent forest types. For example, the boundary between two forest types with similar means but widely different variances should be relatively easy to detect. Similarly, the boundary between two types with the same variance but different means should be easy to identify. With an image generator, one can control both the mean and variance of the simulated forest types. Images will be generated by establishing a template — a thematic map — of known types, assigning a mean and variance to each type for image generation, and generating an image from this information. By having individuals with similar training in the interpretation of aerial photographs examine the images and draw boundaries among types that they identify, we should be able to assess differences in boundary identification and location for various pairs of types relative to their means and variances. From this information we should be able to relate boundary and category errors to the differences in the means and variances of each type.

## Alternative representation for the integration of thematic map and remotely-sensed data

Although not readily apparent, the topics discussed in the previous section provide a conceptual methodology for linking information from satellite images and thematic maps through a common format. In this, the information contained in thematic maps which are stored in a GIS may be considered "vector information" and information stored in satellite images as "raster information." For the purposes of this section, the fashion in which the data is stored internally in the computers is not the issue. That is, it is not our intent to discuss computer algorithms to smooth the "stairstepping" which can occur when raster information is converted to a vector representation. Rather it is our desire to focus on the differences between data which effectively represent a sample grid of points over a region "a raster" and that which draws lines around areas which are considered to be homogeneous at a given map scale "vector polygons".

Currently, in the analysis of remotely-sensed images, considerable activity is devoted to classification procedures with the goal being to produce "better" thematic maps. Although the answer might appear to be obvious, one can ask the question "Why do we strive to do this?" One reason is that when the first remotely-sensed data became available in the early 1970s, it was "understandable" that initial research efforts would be devoted to producing maps that humans could understand. Thus a somewhat implicit goal of image classification was/is to produce the same type of map which is produced from human interpretation of a geographical landscape. As such, the classified image has often been viewed as the final result of a particular analysis. As GIS as a discipline has developed, however, a second reason to produce thematic maps from remotely-sensed data became evident. Because the value of a link between GIS and remote sensing was (and continues to be) recognized, a "common format" was required. Since the goal of many image scientists already was to produce thematic maps (i.e., classified images), it may have seemed "natural" to force the satellite product into a form that would fit into a GIS i.e., a thematic map.

The problem with this latter activity is that by employing only a classified satellite image in a GIS, one has discarded considerable amounts of information. While the goal of a common form of representation remains necessary for suitable data integration, forcing satellite imagery to conform to highly simplified categorical thematic maps means that, potentially, one is operating at an inappropriate level of real-world categorization to perform optimal analysis. If one can move to a less simplified level of data representation, subsequent spatial analysis may prove to be more useful. The most immediate problem is to find a way to be able to represent "GIS thematic map information" at a less simplified, categorized level. That is, one would like to quantify and utilize the raw information that went into making the map.

This is where the error analysis that was discussed in the previous section may be useful. As envisioned, the establishment of a link between boundary and category error should allow one to produce a series of "error" or "certainty" surfaces for a map. One such surface would be required for each class. The surface for a given class would show the certainty or "likelihood" of finding that class at a given location, rather than merely showing a deterministic map having fixed boundaries for all classes.

Such a representation would also be consistent with the classification of remotely-sensed images. All classification techniques are probability-based and the final thematic map that is produced is merely a categorization of probabilities. Usually this is done by assigning a pixel to its "most likely" class with the class to which a pixel is ultimately assigned is merely that class having the highest probability. This may be acceptable if one class's probability is high relative to the probabilties for all other classes. But if the probabilities are nearly equal, and/or if all probabilities are relatively low, considerable information will be lost if a given pixel is merely classed as "Type B".

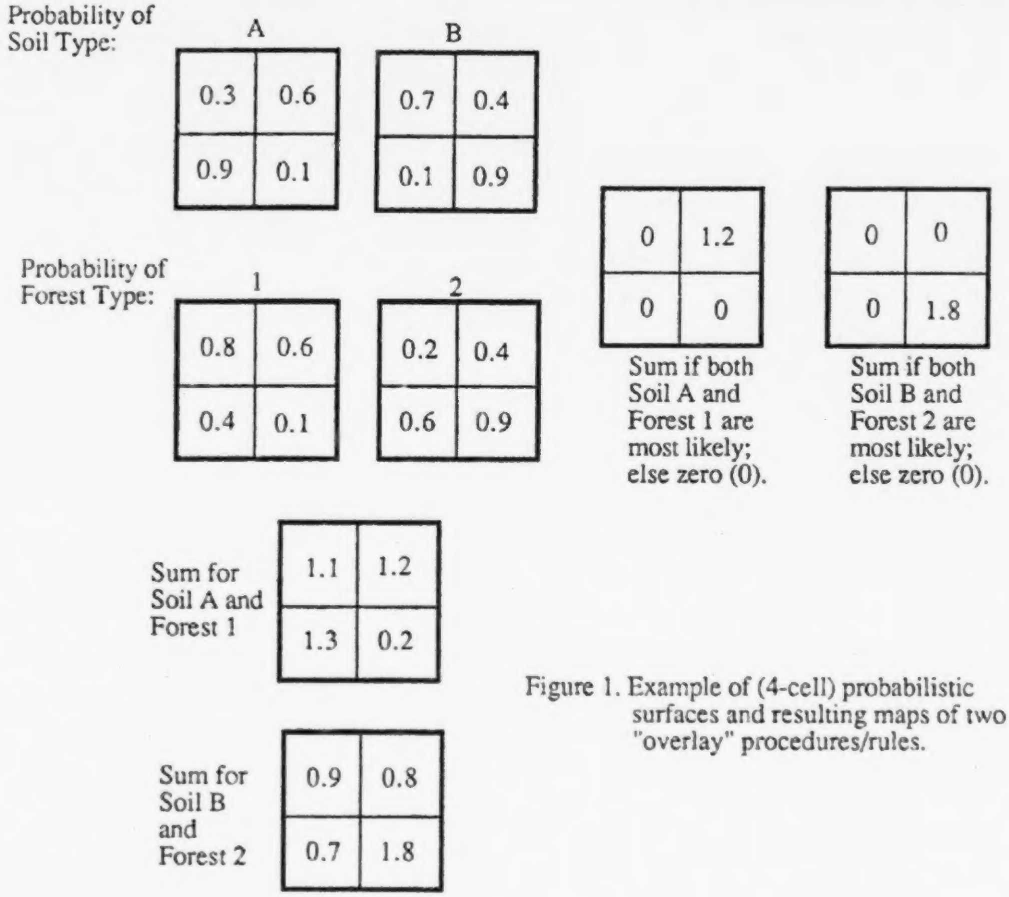
This indicates the potential power of viewing the digital representation of maps as "certainty" surfaces. In deterministic remote sensing analysis, for example, one would classify a pixel as "Type B" if there are 10 possible classes, and the probability of Type B is 0.10 (10%) while the probability for the other 9 classes is equal at 0.09 (9%). (Type B is the "most likely" by virtue of having a higher probability than any of the other classes.) Intuitively, however, this is considerably different from a pixel being classified as Type B because it had a 0.99 (99%) chance of being Type B. Similar dynamics operate on thematic maps. The center of a shallow lake with a highly irregular shoreline has a much higher probability of being "Lake" than those locations at the edge. The work described previously concerning boundary and category errors may allow this and similar concepts to be quantified and incorporated into future spatial analysis. Such a step would provide a valid way for treating "raster" (remotely-sensed) data in the same fashion as "vector" (GIS) data.

## Creation of probabilistic spatial operators

Assuming that one can develop a methodology to create the "certainty" surfaces described and store them in a suitable data structure, it remains to understand how to analyze them. This is particularly difficult to ponder considering the way in which spatial analysis is conducted presently. Current techniques are largely deterministic and based on Boolean logic. A classic example of this type of spatial (and GIS) analysis is that of "factory location." In such a scenario, one wants to determine the "optimal place" at which to locate a factory. In doing so, one must consider the stability of the soil, severity of the terrain, drainage pattern of the region, distance from population centers,...etc. Thus one digitizes a map of each factor, overlays these and identifies in the final step those resulting polygons which have the desired soil-drainage-terrain,...etc. combination. Almost inevitably when this is actually done, those polygons having the desired characteristics are extremely dispersed over the surface and highly irregular in shape and size.

Intuitively, humans understand that one probably cannot actually go into the "real world" and locate the polygons and their exact boundaries. Nor is it the case that these polygons are truly the only "ideal" areas on the map. One must remember that maps of soil, drainage, etc. have variable levels of precision and thus map boundaries are not necessarily located or represented correctly. For example, the real world boundary between two soil types is a "fuzzy" transition zone, not an abrupt boundary as it must be mapped with pen and paper technology. Thus, conceptually, a "more honest" result from the analysis cited would be a map showing the likelihood or "certainty" of having all (or some) of the desired characteristics at a given location.

Thus spatial operators are needed that somehow are based on "certainty" and give answers in terms of the likelihood of finding areas having the desired characteristics. One possibility for analyzing these surfaces would be to use the probabilities stored on the certainty surfaces as weights for each factor. For each location, one might simply sum (or average) the probabilties for each desired class to obtain a score which could be mapped for human interpretation (Figure 1). One potential problem with this is that such scores would mask the fact that the "sufficiency" for one factor might be 1.00, but for another 0.00. Thus an area with a relatively high average score could actually be completely unsuitable. However, techniques could be developed to remove from consideration those areas which have a certainty below a certain threshold for any specified factor. Another possibility for avoiding this problem would be to sum the probabilities only if the desired class on a particular feature map was the most likely.



Probability of Soil Type:

A

| | |
|---|---|
| 0.3 | 0.6 |
| 0.9 | 0.1 |

B

| | |
|---|---|
| 0.7 | 0.4 |
| 0.1 | 0.9 |

Probability of Forest Type:

1

| | |
|---|---|
| 0.8 | 0.6 |
| 0.4 | 0.1 |

2

| | |
|---|---|
| 0.2 | 0.4 |
| 0.6 | 0.9 |

| | |
|---|---|
| 0 | 1.2 |
| 0 | 0 |

Sum if both Soil A and Forest 1 are most likely; else zero (0).

| | |
|---|---|
| 0 | 0 |
| 0 | 1.8 |

Sum if both Soil B and Forest 2 are most likely; else zero (0).

Sum for Soil A and Forest 1

| | |
|---|---|
| 1.1 | 1.2 |
| 1.3 | 0.2 |

Sum for Soil B and Forest 2

| | |
|---|---|
| 0.9 | 0.8 |
| 0.7 | 1.8 |

Figure 1. Example of (4-cell) probabilistic surfaces and resulting maps of two "overlay" procedures/rules.

A conceptually different methodology which would require considerably more rigor and quantitative expertise would be the incorporation of "fuzzy set theory" into GIS analysis. Fuzzy set theory is a mathematical discipline which attempts to account for the idea that not all objects can be classified into one, and only one group. Thus it uses "fuzzy logic" instead of Boolean logic to reach a solution. Of interest here is that one may be able to use "fuzzy information" — i.e., that having some uncertainty associated with it — rather than deterministic data in trying to produce spatial results using fuzzy set theory. This manner of thinking is a significant departure from the way most spatial analyses have been viewed in natural resources. In forestry, for example, it is assumed that a variable of interest can be measured "exactly" except for sampling error (as measured by the statistical standard error). However, the information contained in the standard error is generally not used until "the end" of the analysis. Even at that point, its use is limited largely to the placing of confidence intervals around numerical point estimates. Conceptually, fuzzy set theory differs from this by trying to incorporate "uncertainty" or "error" at every step of analysis.

Regardless of the methodology which is developed ultimately, the next generation of spatial operators also requires an additional change in philosophy. Currently, spatial analysis is done largely by examining the factor at a given location in isolation of objects and factors in the surrounding area. However, the greatest value of maps (arguably) is their ability to show relative location. This unfortunate paradox is evident in most commercial GIS software packages. Though vector systems must internally maintain topology — i.e., information about neighbouring polygons — very few systems allow queries of "adjacency" or "touching." (Fortunately, an increasing number of systems are beginning to be able to do this.) The problem stems partly from the fact that theoretical conceptual models for "neighbourhood analysis" are generally not very well-developed. That is, human understanding of how to define a neighbourhood and what to do with "neighbourhood information" is extremely limted. Yet it seems intuitive that the "value" at any given location will often be strongly related to the "values" at locations "around" it. Current spatial operators work largely using a "pin through the maps" philosophy and examine the characteristics of each location in isolation. Future spatial operators must incorporate the concept of looking across a map surface to evaluate spatial relations rather than down on a location in isolation.

## Data structures

The ideas discussed previously may eventually be found to be extremely difficult to incorporate into existing raster or vector spatial data structures. Both of these structures have a fixed unit which is indivisible — cells for rasters and polygons for vectors. Yet many of the ideas discussed herein require a data structure which has the potential to produce a different "value" for a factor if one moves an "infintessimal" distance across a surface. Furthermore, to be able to develop "across surface" spatial operators, a data structure is required for which an integral part of the data structure is information on the adjacency of objects.

One promising possibility is the Voronoi Diagram which is composed of Theissen polygons. (Terms which are often used instead of, or in relation to, Voronoi Diagrams are "Dirichlait Tesselations" and "Delaunay triangles.") The boundaries of the Theissen polygons for a given map object are defined by the perpendicular bisectors of the object's immediate neighbours (Figure 2) Although one generally constructs Theissen polygons from points, recent work by one of the authors (Gold 1990) has extended the idea to line segments. Each Theissen polygon contains only the area that is closer to a given object than to any other. In doing so, the Voronoi Diagram maintains information about the "neighbours" of every object on the map. Since this information is central to the Voronoi diagram as a data structure, the requirement that adjacency be known and available for all objects is satisfied.

Another requirement that Voronoi diagrams fulfill is that they have the potential to produce a unique value for each point on a map. Gold (1989) has used Voronoi diagrams with "area-stealing" techniques for successful interpolation of topographic surfaces. In "area stealing," one has an existing Voronoi diagram containing Theissen polygons based on (in this case) points of elevation. To obtain the elevation for any given point on the map, one places a "query" point at the desired location and reconstructs the Voronoi diagram as if the point were a "real" point. One then examines the area of the new Theissen polygon of this imaginary point relative to the Theissen polygons on the previous "real" Voronoi diagram. An area-weighted average of the elevations of each Theissen polygon from which area was "stolen" for the new Theissen polygon is then calculated and assigned to the point. The imaginary query point is then removed so that subsequent interpolation will utilize only the original points of known elevation. It can be seen that if one changes the location of the query point even slightly, different areas would be "stolen" from polygons on the "real" Voronoi diagram thus producing a different elevation.

Though continuous (elevation) data were used as an example in the preceding paragraph, a similar concept might be developed for categorical thematic data in order to integrate the topics discussed. One might be able to digitize a thematic map and construct its associated Voronoi diagram. Area stealing techniques might then be used to produce the error surfaces described using the area of each stolen polygon as the certainty. For example, if the insertion of a given point created a polygon which "stole" 60% of its area from a "Type A" Theissen polygon and 40% from a "Type B" Theissen polygon, the certainty of that point being Type A would be 60% with a 40% likelihood of being Type B.

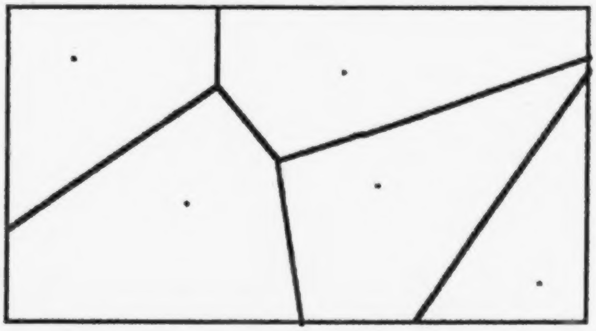


Figure 2. Example of Theissen Polygons.

It is stressed here that such an operation would force one to make a large number of implicit assumptions whose implications have yet to be studied. For example, consider a case where a clearcut is next to a mature forest. Both types have relatively low variances and the boundary between them is fairly readily identifiable. In such a case, one should weight the certainty of a given point being one of the two types by the variance (and possibly mean) of each type. However, an appropriate measure and/or weighting formula for such a concept, and its associated underlying statistical properties have yet to be explored. Furthermore, while area-stealing techniques have proven their value for point data, they have yet to be extended to lines. As noted, the concepts of the Voronoi Diagram have been extended for lines which also allows for the creation of polygons. A similar extension has yet to occur for the area-stealing techniques. Indeed, it is not yet clear that this can be done. By definition, any imaginary "query point" inserted into a polygon-based — as opposed to point-based — Voronoi diagram will steal area only from a single polygon. Thus developing a certainty surface from a polygon-based Voronoi diagram may not be possible, although this is under study.

Nonetheless, the Voronoi diagram as a data structure appears to have the potential to integrate all of the concepts discussed. It has one additional potential advantage that will be mentioned only briefly in closing. Though Voronoi diagrams were initially conceived for point data only, one of the authors has extended these concepts for lines. Thus at its present stage of development, the Voronoi diagram can accomodate both point and line data. If a raster containing, for example, remotely-sensed data can be viewed as a point sample of a surface, it is evident that the Voronoi diagram can readily accomodate this information. Similarly, if one has a thematic soils map with linear boundaries, one can also incorporate this into a Voronoi diagram. Thus the ability to handle both point and line data in the same fashion may provide additional benefits from the Voronoi diagram as a data structure.

## Conclusion

As DSTs evolve and progress, four issues must be addressed. First, there must be better ways to quantify and utilize spatial errors. In particular, "real-world-to-map" simplifications must be identified, measured, and incorporated into spatial analysis. Second, the use of these errors should provide a natural way to integrate remotely-sensed and thematic map-based data. Third, the integration of the two data types into a single structure will require a new family of spatial operators for spatial analysis. Rather than operating on the classified thematic maps in the GIS, these operators should utilize the underlying error surfaces. They should also work "across" maps rather than employing a "pin through the map" concept. Finally, the embodiment of these concepts into usable software algorithms will probably require the development of a new spatial data structure. Such a structure will require that each point on the map has the potential to be assigned a different value, and that information about adjacent map objects be an integral part of the structure. The Voronoi diagram has already shown potential to meet both of these needs.

## Literature cited

Chrisman, N.R., 1982. Methods of spatial analysis based on error in categorical maps. Unpublished Ph.D. Thesis, University of Bristol (England). 261 pp.

Gold, C., 1989. Surface interpolation, spatial adjacency, and GIS. In: *Three-dimensional Applications in Geographical Information Systems*, J.R. Raper (ed.) (London: Taylor and Francis), p. 21-35.

Gold, C., 1990. Neighbours, adjacency, and theft the Voronoi process for spatial analysis. Proceedings: *First European Conference on Geographic Information Systems.* Amsterdam, Netherlands, April, p. 382-391.

Robinson, V.B., 1988. Some implications of fuzzy set theory applied to geographic databases. Computers, Environment, and Urban Systems 12:89-98.

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Spatial considerations for forest canopy modelling and biomass estimation based on remotely sensed data

**Geoffrey Edwards**
**Kim Lowell**

Industrial Chair in Geomatics Applied to Forestry
Centre de recherche en géomatique
Faculté de foresterie et géomatique
Pavillon Casault, Université Laval
Sainte-Foy, Québec, G1K 7P4

## Abstract

*Considerable effort is being undertaken presently to model forest canopies, both for microwave imaging and high resolution VIR data. The reasons for this effort are many fold, including the study of bidirection reflectance, the evaluation of stress conditions in trees, forest inventory tasks and multi-temporal studies of evolving, seasonal or diurnal forest conditions. One of the principal reasons is that it is believed that such modelling may yield a means of estimating forest biomass, especially from microwave imaging at low frequencies. Ground truth acquisition has focussed on the collection of physical parameters such as the dry density of canopy components, their dielectric constants, etc. and, more recently, on canopy architechture data. This is partially in response to the increasing sophistication of the models. However, sampling practices tend to focus on traditional stand parameters or on managed forests (plantations). In this paper, we examine the spatial variability and spatial structure of forest stands with a view to identifying which model parameters are constant over the forest stand. We find that although the stand is highly variable, there may exist an intermediate scale structure over which the gross canopy characteristics are more homogeneous. The existence of these intermediate scale patches would render possible a reduction in the number of degrees of freedom of spatial canopy models. Furthermore, biomass estimates based on these regions might be more reliable than those derived from the traditional forest stand. More work needs to be carried out to study such within-stand structure.*

## Introduction

Many different kinds of forest-based models are currently under development. In fact, the 1990s is probably going to be the "decade of the model". These models include yield prediction models under development by government and industry to help plan cutting operations, bidirectional reflectance (BRDF) models used to help understand sun-vegetation-sensor view-angle geometries for remote sensing images (Strahler and Jupp, 1991; Fournier and Miller, 1989), radar backscatter models used to help understand scattering processes between canopy components and microwave signals (Dobson et al, 1991) and, particularly, with a view to assessing biomass directly (Sader, 1987), forest ecological models used to understand chemical cycling in the canopy, process models used to understand gas exchange processes between the canopy and the atmosphere (McElroy and Wofsy, 1986) or the effects of changing spectral properties of the canopy on near-surface climatology (Sato et al, 1989), hydrological models used to understand water uptake in the canopy

(Roberge and Plamondon, 1987) and so on.

The majority of these models are strictly attribute models with no spatial component. They use a number of intrants, varying from a few to several hundreds and produce a number of extrants. Some of the models are physically-based, attempting to model the forest environment based on biophysical understanding of its constituent components. Other models are purely empirical, profitting from observed correlations between different parameters of interest. Physically-based models have the advantage that many of the parameters are circumscribed by physically realistic domains, whereas purely empirical models are hampered by excessive degrees of freedom. However, whatever the model's structure, they all suffer from similar problems.

These complex models usually have many more degrees of freedom than can be fixed by a single set of data. Ground data must be acquired under many different conditions (and also several times under similar conditions) in order to properly parameterize and validate these models. Frequently, the model parameters are not directly measurable but must be inferred from measured values. Furthermore, the model may require parameters to be fixed which are not measurable at all, but represent intermediate conditions in the modelled behavior.

Much of the modelling process which has gone on to date has involved setting up a number of ground sites at which extensive measurements are made. Each ground site is dealt with as if it were independent of the other sites.

A great deal of interest is presently being focussed on the use of remotely sensed data to provide inputs to some of these models. The largest obstacle to doing so appears to be that remotely sensed data is acquired by averaging over some surface area, varying between ten meters squared and several kilometers squared for satellite sensors, whereas ground data is acquired at an individual "point". The problem then is to "regionalize" the point measurements so that they can be compared with remotely sensed "pixels".

Remotely sensed data offers some major advantages for forest modelling. The high number of "free parameters" of many of these models can be reduced dramatically by using contiguous pixels which are known or can be assumed to share common behavior. The problem then becomes to determine the spatial scales at which different model parameters can be assumed to be constant or slowly varying. Many of the forest models discussed in the introduction are CPU-intensive, and hence it is of the utmost importance to determine the correct "sampling" scales at which they should be run to compare with image data as well as the scales over which ground measurements should be made for validation purposes.

Much of the ground truth data presently being targetted for the remote sensing related modelling (BRDF, radar, process models, etc.) involves physical parameters such as the dry density of canopy components, their dielectric constants, the geometrical architecture of canopy constituents, their optical properties, etc. However, much of this work is being done on the assumption that the stand is the basic unit of interest. Sampling schemes adopted within stands are carried out on a random basis in order to characterize their statistical variability. The same is true of yield prediction models.

## The regionalization of forest parameters

To test this hypothesis, we carried out a sampling campaign in the Montmorency Forest, an experimental forest operated by the Faculty of Forestry and Geomatics of Laval University in the Laurentian Mountains north of Québec City. The sampling campaign was actually designed around two goals: (a) a study of stand boundary errors derived from photo-interpreted aerial photos, and (b) a study of spatial variability for vegetation modelling. The study region is situated in Figure 1 with the sampling pattern laid down. The sampling pattern adopted consisted of an 8 by 8 grid of sample points placed every fifty meters. The grid passes through at least five different stands. Some pairs of adjacent stands are quite different, others are very similar.

The Montmorency forest is characterised by balsam fir stands and mixed stands whose deciduous component is dominated by birch. The area has moderately high relief. Several areas of blowdown exist within the sector sampled. These areas are due to the spruce budworm infestation which occurred over the previous decade.

## Analysis of the grid data

The grid point locations are shown in Figure 1(b). The grid sample points were located using GPS positioning techniques, involving the placing of a 25 foot mast at three corners of the grid and a fourth GPS point at the intersection between two trails which are visible in the images. GPS locational accuracy was around 50 cm, whereas internal geometrical accuracy of the image registration (obtained from analysis of residuals) is of the order of 2 meters. However, additional error in position also seems to be present, probably due to effects of relief on the positioning, leading to a present accuracy not much better than 10 meters. This will probably be reduced in the future as our correction techniques are modified to account for variable topography, but for the purposes of this initial experiment, this accuracy was all that was available.

At each grid location, the following measurements were taken: the height, dbh, species, crown width and crown depth of each of three characteristic trees, the number of stems in each two centimeter dbh class from 0 to 10 centimeters and each one centimeter class above 10 centimeters, for each species. Also, a visual estimate of the slope and aspect, local stand characteristics

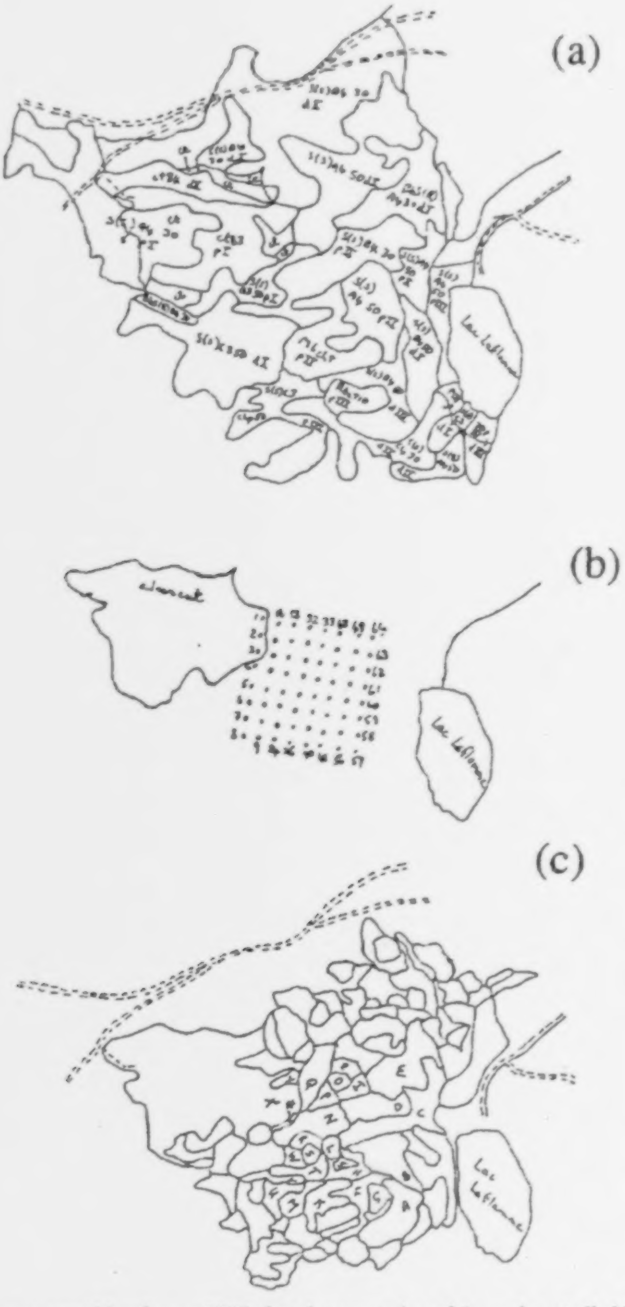

(a)
Lac Laflamme
(b)
clearcut
Lac Laflamme
(c)
Lac Laflamme

**Figure 1:** (a) The forest map resulting from a 1985 photointerpretation of the region studied in this project; (b) the location of the grid points with respect to the major features of the region; (c) the result of the 1989 photointerpretation of the region, using a finer interpretation key than is usual. The 27 regions which enclose one or more sample plots are labelled with letters.

(density and height classes and dominant species), blow-down if present and low growth was made at each grid location, as well as an estimate of relative crown closure in each of four quadrants. From these data, the basal area was determined for each sample plot. The spatial variability of these data is high, as expected. Uniformity within stands seems very hard to obtain. However, several pairs of neighbouring sample locations exhibit similar behaviour in one or two of the inventory variables. Visual observations made at the sites confirm this observation, and, in addition, these sites fall in the same fine-scale polygons derived from the interpretation of the aerial photo data (see next section). This interpretation was carried out before any reliable positioning of the sample points onto the image was obtained.

## Analysis of the aerial photo data

A stereo pair of 1989 aerial photos of the same region was studied and interpreted using a key tuned to finer distinctions than that which is usually used to produced forest inventory maps. The reason for this is that we believe we can more easily produce automated classifications of forest units at such finer scales using high resolution digital imagery than the difficult if not impossible task of obtaining forest stand boundaries similar to those usually obtained by expert human interpreters. By looking for relatively small homogeneous units, we were able to interpret the forest image for the region covering the sample grid. The results of this interpretation are shown in Figure 1(c), and can be compared with the stand boundaries for the region obtained from a standard forest map (Figure 1(a)).

The grid points were overlaid onto both the forest inventory map dating from 1985 and the new interpretation. The new interpretation fits the old forest map reasonably well, considering the errors intrinsic to the photointerpretation process. Each forest stand contains between two and seven of the new regions, with a few shifts in the boundaries.

The grid points could not be located to better than about ten meters accuracy. For the purposes of this experiment, the descriptive data obtained during the ground inventory were used to help locate individual points as best as possible. There is still likely to be positioning errors above five meters and hence some of the grid points may not fall in the regions to which they have been allocated, but rather into some nearby neighbour.

For the analysis, the sets of neighbours which were likely candidates for such positional confusion were tabulated. The initially assigned positions were used for the subsequent processing, but they were also perturbed to deal with confusion neighbours, to get some feeling for how the positional errors might affect the results.

The analysis consisted of studying the within-group and between-group variances of a variety of inventory parameters for the forest stands obtained from the 1985 photointerpretation and comparing these values with those obtained with the new, finer interpretation. In addition, tests were run to determine the significance of the results by dividing the 1985 stands into roughly the same number of groups as found in the 1989 interpretation on a purely random basis, rather than spatially partitionning the data.

**Table 1: 1985 Forest stands which contain grid sample points and their associated 1989 regions**

| Forest stand | # Stand label 1985 stands | Polygon label 1989 regions |
|---|---|---|
| 1 | S(S)A4 50 | A,B,C,F |
| 2 | S(S)B4 30 | D,E,I,N,R,T,W |
| 3 | M6 Cht | G,H,J,K,L,S |
| 4 | ct | X,Y |
| 5 | S(S)C3 50 | M,U,V |
| 6 | BbS(R)A4 30 | O,Z |
| 7 | S(S)C3 50 | P,Q |
| 8 | S(S)B3 50 | XX |

The new interpretation divided up the original grid of 64 sample points into 27 new regions, compared to the 8 forest stands obtained from the 1985 interpretation (see Table 1). Stand #4 was excluded from the analysis because it consists of a clearcut, and stand #7 was also excluded because the two sample points each fall in single regions (P and Q), both of which are characterised by severe blow-down. Several of the new regions contained a single grid point. Statistically, these polygons do not contribute at all to the within-group variance, because no variance can be determined on a single data point. Nonetheless, the contribution of these points to the observed increase in between-group variance was generally less than a third. (See Table 2)

It was found that the between-group variance of basal area increased from 81 square centimeters to 1315 out of a total variance of 4176, between the old photointerpretation and the new interpretation scheme (Table 2). If the results within a stand were randomly distributed in a statistical sense, one would expect no significant change in the between-group variance upon subdividing the region further. The analysis of several random divisions of the data implied a total increase in between-group variance to about 600 square centimeters, or half that found with the spatial partition. Most of the increased between-group variance comes from stand 2, where several small regions which contrast sharply in appearance with their neighbours are found. The other stands did not appear to show significantly increased separability in basal area.

**Table 2: Increase in between-group (bg) variance for basal area between traditional forest stands and the 1989 finer spatial partition.**

| Forest stand | BG Variance 1985 stands | BG Variance 1989 regions | Increase in BG Variance (from stands to regions) |
|---|---|---|---|
| 1 | 8.78 | 133.67 | 124.89 |
| 2 | 0.06 | 862.25 | 862.19 |
| 3 | 37.45 | 156.70 | 119.25 |
| 5 | 29.44 | 157.44 | 128.00 |
| 6 | 0.70 | 0.70 | 0.00 |
| 8 | 4.68 | 4.68 | 0.00 |
| TOTAL | 81.11 | 1315.44 | 1234.33 |

**Table 3: Increase in between-group (bg) variance for number of regenerating trees between traditional forest stands and the 1989 finer spatial partition.**

| Forest stand | BG Variance 1985 stands | BG Variance 1989 regions | Increase in BG Variance (from stands to regions) |
|---|---|---|---|
| 1 | 1 | 1393 | 1392 |
| 2 | 40 | 1974 | 1934 |
| 3 | 1 | 2472 | 2472 |
| 5 | 414 | 619 | 205 |
| 6 | 268 | 268 | 0 |
| 8 | 44 | 44 | 0 |
| TOTAL | 767 | 6770 | 6003 |

On the other hand, the between-group variance increased much more for the number (density) of regenerating trees under the canopy. The random division of the stands led to an increase in between-group variance of around 2000, whereas the spatial partition of the data increased between-group variance to 6770 (out of a total variance of 11853). Of this, 3400 was contributed by polygons containing a single data point. Closer examination of the source of this increase revealed that stands 1, 2 and 3 all contain small zones with quite different characteristics (and appearance) which are well separated by density of regenerating trees (by definition, trees with diameter smaller than 10 centimeters). However, it is also clear that these regions do not account fully for the spatial partition of the finer intrepretation. Some other indicator must also be present to account for these differences.

The between-group variance stayed at the same level for stands 5 (S(S)C3 50), 6 (BbS(R)A4 30) and 8 (S(S)B3 50). All three are characterised by relatively subtle differences within the stands. Stands 1, 2 and 3 however, all included several more visible effects in the photo. Stand 7 (S(S)C3 50) was excluded from the analysis, because both data points were in recent blow-down ("chablis") containing no standing timber and hence no equivalent measurements as those obtained for the other sample plots. The increase in between-group variance for stand 1 is almost all due to the presence of significantly different values for the F region — the rest of the stand is fairly uniform. (The breakdown of variance by its 1989 region is shown in Table 4.) Stand 2 (S(S)B4 30) has several clearly defined regions in the photo and the statistically significant increase in between group variance by dividing it up supports this. Stand 3 (M6 cht) also experiences large variations, as would be expected in an old blow-down ("chablis total" even in 1985). The analysis of random divisions of this data indicate that the increase in between-group variance for these stands *is* significant, if marginally so (between two and three standard deviations from random).

Finally, the between-group variance for the crown to total height ratios (C/H) shows important variations for some of the 1989 regions (especially those within stands 2, 3 and 5, shown in Table 5). This parameter is important because the radar backscatter is very sensitive to the crown to total height ratio.

**Table 4: Breakdown of between-group (bg), within-group (wg) and total variance for the number of regenerating trees in the 1989 regions.**

| Forest stand # 1985 | Region # 1989 | N Variance | BG Variance | WG Variance | Total |
|---|---|---|---|---|---|
| 1 | A | 2 | 12 | 13 | 65 |
| 1 | B | 8 | 154 | 578 | 732 |
| 1 | C | 3 | 303 | 41 | 344 |
| 2 | D | 7 | 28 | 1198 | 1224 |
| 2 | E | 3 | 174 | 650 | 824 |
| 1 | F | 4 | 884 | 1697 | 2581 |
| 3 | G | 1 | 135 | 0 | 135 |
| 3 | H | 1 | 268 | 0 | 268 |
| 2 | I | 3 | 429 | 129 | 558 |
| 3 | J | 1 | 237 | 0 | 237 |
| 3 | K | 4 | 46 | 174 | 220 |
| 3 | L | 1 | 1269 | 0 | 1269 |
| 2 | N | 5 | 564 | 590 | 1154 |
| 6 | O | 1 | 268 | 0 | 268 |
| 2 | R | 2 | 2 | 613 | 615 |
| 2 | T | 3 | 411 | 83 | 494 |
| 5 | U | 1 | 207 | 0 | 207 |
| 5 | V | 1 | 207 | 0 | 207 |
| 2 | W | 3 | 366 | 43 | 409 |
| 8 | XX | 1 | 44 | 0 | 44 |
| TOTAL | | | 6047 | 5806 | 11853 |
| TOTAL (excluding single points) | | | 3412 | 5806 | |

**Table 5: Increase in between-group (bg) variance for crown to total height ratio (c/h) between traditional forest stands and the 1989 finer spatial partition.**

| Forest stand | BG Variance 1985 stands | BG Variance 1989 regions | Increase in BG Variance (from stands to regions) |
|---|---|---|---|
| 1 | 0.13 | 0.15 | 0.02 |
| 2 | 0.02 | 0.18 | 0.16 |
| 3 | 0.00 | 0.51 | 0.51 |
| 5 | 0.06 | 0.22 | 0.15 |
| 6 | 0.10 | 0.09 | 0.00 |
| 8 | 0.09 | 0.09 | 0.00 |
| Total | 0.41 | 1.24 | 0.83 |

It should also be emphasised that the decision criteria used were not based on the ground inventory data at all, but rather on the appearance of a 1989 stereo aerial photo pair. The partitioning of the data occurred spatially and not in the parameter space (only the random breakdown of the individual stand data were based on parameter space).

The differences in the ground inventory data found between the two spatial partitions are statistically significant, but are nonetheless fairly subtle - they are difficult to pick out of the inventory data on the basis of a cursory examination. On the ground, the effects appeared to be more striking, but we are unsure what to measure to reinforce this subjective impression. The new interpretation of the aerial photo seems to correspond reasonably well to the percieved differences on the ground, since the between-group variance is significantly improved

## Discussion and conclusions

Hence the analysis showed that the between-group variance was sigificantly higher for many of the inventoried parameters when comparing the traditional photointerpretation with a new, fine-scale photointerpretation. We were unable to account for more than half of this increased variance by dividing the traditional forest stands into similar numbers of statistically random groups. The increased between-group variance in "surface terrière" could be attributed to the strong contrast differences found in stand 2, but several other measurements, including the number of regenerating trees underneath the canopy and the crown to total tree height ratio accounted for the local homogeniety found in some of the other stands, particularly stands 1 and 3. Although the difference in dates (1985 versus 1989) may account for some of this change, the affected regions in the 1989 data are too small to be identified as standard forest stands.

Nonetheless, the observed between-group variance is not enough to fully account for either the homogeniety in the small regions obtained from the 1989 photointerpretation or the differences between these regions. Furthermore, it is not clear that either the inventoried variables or the small regions obtained from the 1989 photointerpretation relate to the small-scale homogeneity observed on the ground. Are we looking at three different spatial entities or a single kind of entity from three different perspectives? The sample spacing was too coarse to properly evaluate these small regions, since in the best case only 6 or 7 data points were located in a single region and most of the time three or less data points were present. Furthermore, it is suspected that the variables inventoried in the ground data do not fully measure the small-scale homogeniety.

The small regions characterised by high contrast differences with surrounding areas may be important for standard forest inventory, since they do seem to significantly affect the effective timber volumes determined from the ground inventory data. The small regions which are less contrasted with their neighbours may not be useful for forest management but are likely to be important for certain kinds of modelling work, since some of their characteristics seem to be accounted for by variables such as regeneration under the canopy and crown to total height ratios. The latter is known to be very important for radar backscatter modelling at almost all frequencies, and the former is probably important for modelling the scattering of light inside the forest canopy in bidirectional reflectance modelling. Other variables may also be at work, although their nature is not yet clear.

In addition, these finer scale polygons are relatively easy to identify in stereo photo pairs and preliminary work (in preparation) indicates that they may be easier to obtain by automated techniques from high resolution airborne imagery than traditional forest stands.

These smaller regions may simply express more detailed forest inventory information, similar to that used in several European countries such as Switzerland (Schmid-Haas, 1976). These inventory systems sometimes include information on lower stages within the forest canopy as well as small spatial groupings and mixtures of trees at local scales. Furthermore, extensive work on ecology-based forest inventory may account for some of the fine scale structure seen and measured here. The results of this analysis, although not conclusive, do indicate that more work is warranted to study the relationship between what is observed at small scales (within the traditional forest stand) from airborne sensors and what is measured on the ground.

This is especially important for modelling efforts in remote sensing, since our ground observations (admittedly subjective) seem to indicate that many tree architectural parameters may be relatively stable inside the small units studied here. For the modelling of both radar backscatter and the bidirectional reflectance distribution of visible light off the forest canopy, these considerations are likely to be important. Furthermore, the fact that these smaller units may be easier to extract from remotely sensed imagery may make them useful for forest management strategies, even though they are at finer scales than those used by existing techniques.

## Acknowledgements

We would like to thank the Association des Industries Forestières du Québec as well as the Natural Sciences and Engineering Research Council of Canada for the funding necessary to do this work.

## References

Dobson, M.C., K. McDonald, F.T. Ulaby and Terry Sharik, 1991, "Relating the Temporal Change Observed by AIRSAR to Surface and Canopy Properties of Mixed Conifer and Hardwood Forests of Northern Michigan", *Jet Propulsion Laboratory SAR Workshop*, Pasadena, California.

Fournier, R.A. and J.R. Miller, 1989, "A 3-dimensional modelling of forest canopies for high resolution imagery", *Proceedings of the IGARSS'89 Symposium*, Vancouver, B.C., 2370-2373.

McElroy, M.B. and S.C. Wofsy, 1986, "Tropical Forests: Interactions with the Atmosphere" in *Tropical Rain Forests and the World Atmosphere*, G.T. Prance (ed.), Westview Press, Boulder, Colarado, 33-60.

Roberge, J. and A.P. Plamondon, 1987, "Snowmelt Runoff Pathways in a Boreal Forest Hillslope, the Role of Pipe Throughflow", *Journal of Hydrology*, Vol. 95, 39-54.

Sader, S.A., 1987, "Forest Biomass, Canopy Structure, and Species Composition Relationships with Multipolarization L-Band Synthetic Aperture Radar Data", *Photogrammetric Engineering and Remote Sensing*, Vol. 53, 193-202.

Sato, N., P.J. Sellers, D.A. Randall, E.K. Schneider, J. Shukla, J.L. Kinter III, Y.-T. Hou and E. Ablertazzi, 1989, "Effects of implementing the Simple Biosphere Model (SiB) in a general circulation model", *Journal of Atmospheric Science*, Vol. 46, 2757-2782.

Schmid-Haas, P., 1976, "Aufgabe und Anwendung der Bestandeskarte", *Eidg. Anst. forstl. vers'wes.*, Vol. 147, 7-18.

Strahler, A.H. and D.L.B. Jupp, 1991, "Modelling bidirectional reflectance of forests and woodlands using boolean models and geometrical optics", *Remote Sensing of Environment*, in press.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Regeneration stocking derived from high resolution airborne imagery via texture measurements

**Edwards, G.**

Chaire industrielle en géomatique appliquée à la foresterie
Centre de recherche en géomatique
K.P.B. Thomson and N. Eldridge
Centre de recherche en géomatique
Faculté de foresterie et géomatique
Pavillon Casault, Université Laval
Ste-Foy, Québec, G1K 7P4
Fax: (418) 656-7411

## Abstract

*Image texture has been proposed by many investigators to supplement standard spectral classifiers and improve classification results. The benefits are expected to be especially high in high resolution airborne data where traditional spectral classifiers do poorly because of the high variability of the data. An alternative use of texture is to look for direct correlations between texture measures and parameters of interest. In this paper, results of using this approach to studying forest regeneration after clear cutting are examined. Percent stocking for young trees higher than one meter is found to correlate directly to three texture measurements, despite the presence of competing vegetation. These texture measurements are found to provide different information than that contained in the spectral bands. It is hence feasable to evaluate stocking automatically from digital analysis of high resolution airborne imagery.*

## Introduction

The inventory of forest regeneration plays an essential role in competent long-term forest management. If obtained early after harvesting, it can aid in selecting the most efficient application of sylvicultural techniques and thus enhance the productivity of the site. Such evaluation is presently carried out using site visits supplemented by photo-interpretation of aerial photos. With the advent of high spatial resolution digital imagery such as the MEIS II sensor, it might be possible to develop more quantitative techniques for evaluating regeneration conditions in these images. Earlier studies of the MEIS II data indicated some potential for separating different levels of regeneration (Labonté et al, 1987).

An experiment was undertaken to study the relative information portrayed by textural versus spectral measurements of regeneration and the effects of competing vegetation on these measurements.

## The data

The MEIS II sensor was developed for the Canadian Centre for Remote Sensing by McDonald Dettwiler. It is a linear array scanner of the "pushbroom" type with both high spatial and spectral resolution. The

TABLE 1: STOCKING CLASSES FOR REGERATION TALLER THAN 1 METER

| Region | Percent stocking |
|---|---|
| A | 60 -100 |
| B | 40 - 60 |
| C | 0 - 40 |

TABLE 2: TEXTURE MEASUREMENTS FOR EACH STOCKING CLASS. SHOWN ARE MEAN VALUE, STANDARD DEVIATION ("SIGMA") OF THE SAMPLES AND UNCERTAINTY ("ERROR") OF THE MEAN. THE USE OF THE EXPRESSION "(E2)" INDICATES THAT THE ASSOCIATED VALUES SHOULD BE MULTIPLIED BY 100

| | Class A | | | Class B | | | Class C | | |
|---|---|---|---|---|---|---|---|---|---|
| Band Measure | Mean | Sigma | Error | Mean | Sigma | Error | Mean | Sigma | Error |
| 2 Intensity | 88.00 | 19.00 | 0.26 | 101.00 | 20.00 | 0.28 | 100.00 | 14.00 | 0.19 |
| 7 Intensity | 92.00 | 13.00 | 0.18 | 95.00 | 11.00 | 0.15 | 91.00 | 7.00 | 0.10 |
| 3 Intensity | 88.00 | 12.00 | 0.17 | 80.00 | 11.00 | 0.15 | 73.00 | 8.00 | 0.11 |
| 2 Contrast | 10.96 | 1.65 | 0.21 | 10.57 | 1.66 | 0.21 | 8.85 | 1.50 | 0.19 |
| 7 Contrast | 7.13 | 1.52 | 0.19 | 5.30 | 1.45 | 0.18 | 4.47 | 0.85 | 0.11 |
| 3 Contrast | 8.88 | 2.04 | 0.26 | 6.55 | 1.75 | 0.22 | 5.14 | 0.93 | 0.12 |
| 2 Energy | 0.008 | 0.001 | 0.000 | 0.008 | 0.001 | 0.000 | 0.009 | 0.001 | 0.000 |
| 7 Energy | 0.009 | 0.001 | 0.000 | 0.011 | 0.002 | 0.000 | 0.012 | 0.001 | 0.000 |
| 3 Energy | 0.009 | 0.001 | 0.000 | 0.010 | 0.001 | 0.000 | 0.011 | 0.001 | 0.000 |
| 2 Entropy | 2.092 | 0.015 | 0.002 | 2.092 | 0.016 | 0.002 | 2.074 | 0.023 | 0.003 |
| 7 Entropy | 1.788 | 0.018 | 0.002 | 2.011 | 0.047 | 0.006 | 1.986 | 0.047 | 0.006 |
| 3 Entropy | 2.068 | 0.028 | 0.004 | 2.039 | 0.039 | 0.005 | 2.009 | 0.040 | 0.005 |
| 2 Cl. shade (E2) | 13.55 | 83.61 | 10.45 | -1.71 | 91.18 | 11.40 | -4.04 | 93.81 | 11.73 |
| 7 Cl. shade (E2) | 4.10 | 30.30 | 3.79 | -1.38 | 38.17 | 4.77 | -6.71 | 25.95 | 3.24 |
| 3 Cl. shade (E2) | 4.61 | 32.58 | 4.07 | 8.60 | 44.33 | 5.54 | -2.02 | 21.60 | 2.70 |
| 2 Cl. prom. (E5) | 11.40 | 10.00 | 1.25 | 11.60 | 15.10 | 1.89 | 7.20 | 13.30 | 1.66 |
| 7 Cl. prom. (E5) | 2.94 | 2.70 | 0.34 | 1.94 | 4.24 | 0.53 | 0.86 | 1.81 | 0.23 |
| 3 Cl. prom. (E5) | 3.51 | 3.34 | 0.42 | 2.35 | 4.56 | 0.57 | 1.08 | 2.02 | 0.25 |

system may contain up to eight filters in the visible and near infrared. Details can be found in Till et al (1983).

The image data were acquired in September 1986 over the Montmorency Forest, an experimental boreal forest site in the Laurentian Mountains belonging to Laval University and located less than an hour's drive from the campus. Although imagery at three different resolutions (altitutes) were acquired, only the highest resolution (70 cm.) data were used for this study. At the time the MEIS data were acquired, the competing vegetation was largely dormant. The raspberry had lost all its leaves and the herbaceous plants their chlorophyl.

A single large clearcut which contained three different sets of regeneration conditions was identified and inventoried in September 1989 (Brunelle, 1991). Eighty four sample plots were established in a regular grid pattern over about eight hectares (the layout of the site is shown in Figure 1). For each sample point, the number of regenerating conifer seedlings within a 1.13 meter radius was determined along with their species identifications and heights. Furthermore, qualitative estimates of the density and height of competing herbaceous and wild raspberry over a 10 meter radius was obtained, along with an estimate of the density and height of the regeneration, and a one word description of the topography and drainage conditions of the site. The slopes were generally low to medium, since this site was picked because it is relatively flat, and moderate to good drainage conditions were reported.

The regeneration was observed to consist principally of balsam fir, with a few black spruce seedlings in regions less well drained. Seven height classes were determined (from 0 to greater than 3 meters) for estimating both the regeneration and the competing vegetation heights over the ten meter radius samples.

For the correlations, the height of the conifer stems was "backprojected" three years to obtain an estimate of the regeneration height at the time of the overflight. This involved "shrinking" the stems by about one meter. Even after such extrapolation, which is believed to be overestimated, the ratio of stem height to competing vegetation is at least a factor of two. Furthermore, the competing vegetation was not "backprojected" to 1986 and hence was also probably overestimated.

The area was cut in 1975 and then again in 1978 (Figure 1). In addition, good drainage conditions on the east end of the clearcut led to better regeneration than on the west side. Three sub-regions were selected for the texture analysis, one on the east end which was relatively well regenerated in 1989 (class A), a second poorly regenerated region at the west end (class C) and an intemediate region, both spatially and in terms of regeneration (class B). These classes were not defined purely on the basis of spatial location, however, but also on the density (percent stocking) of seedlings more than one meter in height for the smaller radius sample plots (Table 1), after backprojecting the measurements from 1989 to 1986.

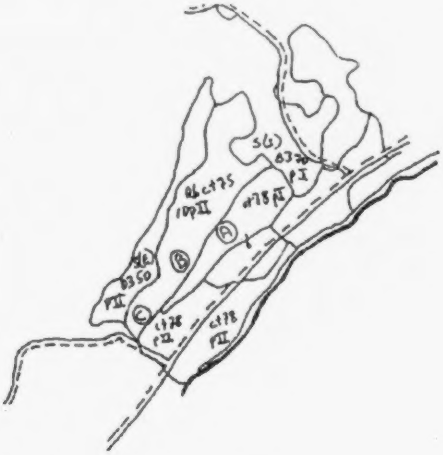


Figure 1: A part of the 1985 forest map for the cutover region. The heavy line is a stream, the double lines are dirt roads and the three regions used for the study are indicated by the circled letters A, B and C

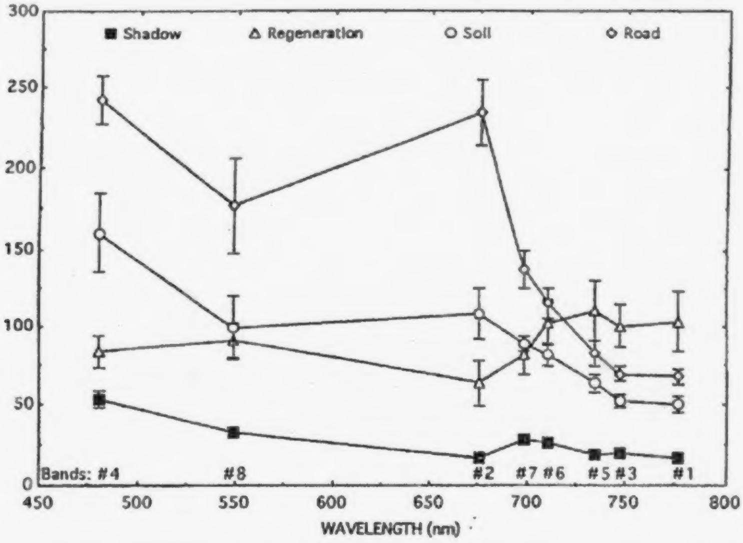


Figure 2: Grey levels for each of four cover types as a function of wavelength. Each value is an average over about twenty pixels, hence the error bars.

Three spectral bands out of the eight available were selected for the study (#2 at 675±40 nm., #7 at 698±13 nm. and #3 at 747±17 nm.). These were chosen by studying the relative intensities of different ground types in different bands (Figure 2). As representative a set of pixels as was possible was chosen for each ground type and used to derive the different spectra. The class called "soil" includes the competing vegetation, whereas the class "conifer" probably includes some contribution from shadowing. Bands 2 and 3 were selected because they maximized the contrast between pixels which were mostly conifer regeneration and pixels which were mostly soil, although the relative intensities of the two ground types are inverted between these two bands (vegetation is bright in band 3 but dark in #2). Band 7 was chosen because it is situated at the crossover point, where conifer pixel intensities equal soil pixel intensities, although it maintains a moderately high contrast for vegetation versus shadow. It was felt than band 7 might serve as a control set of data.

## The analysis

The ground measurements were matched to the image data through careful measurements with respect to easily visible features in the image, and superimposed on the image via the PAMAP Geographic Information System, which was used to manage the data. Texture measurements within 9 x 9 windows were extracted throughout the region where the sample plots were located. The window size was determined by earlier work as being optimal for balancing a reasonable size statistical sample of pixel values against decreased spatial resolution (Edwards et al, 1988), and also because at 70 centimeters spatial resolution, this corresponds to a region of similar size as that of the sample plot separation.

The mean intensities in each of the three bands for each of the three sectors are shown in the first three lines of Table 2. Although some differences are found between the three sectors, the effects are not very strong. It is interesting to note that the contrast effects expected based on the band selection process are present in these data. The low intensities in band 2 are associated with class A as expected, since the conifer regeneration is darker than the soil class, whereas they are associated with class C in band 3 where the soil is darker than the regeneration. Furthermore, band 7 shows no discrimination, as expected because of its location at the crossover point. The histograms for each of the three bands are shown in Figure 3.

The fact that a weak correlation is found in the purely spectral data supports results found earlier (Labonté et al, 1987; Kneppeck and Ahern, 1987).

The texture measurements used were cooccurence measures. These are second order statistical measures. They are derived from the so-called cooccurence matrix (also called the grey level dependance matrix or GLDM), which is a matrix which records the probability density functions for finding a pixel of grey level j at a distance d and an angle q from a pixel of grey level i. In practice, only four angles are considered, 0°, 45°, 90° and 135°

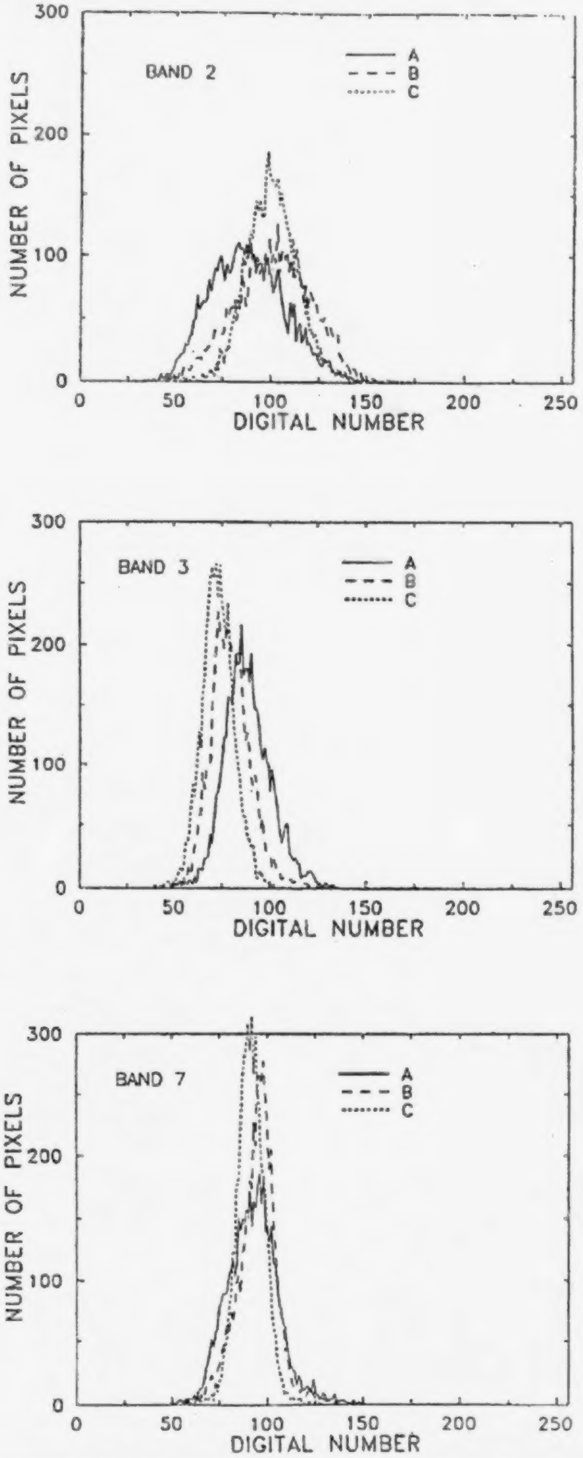


Figure 3: Grey level histograms for each of the three bands used in the study and each of the three regeneration classes. Class A is the most regenerated class, Class C the least.

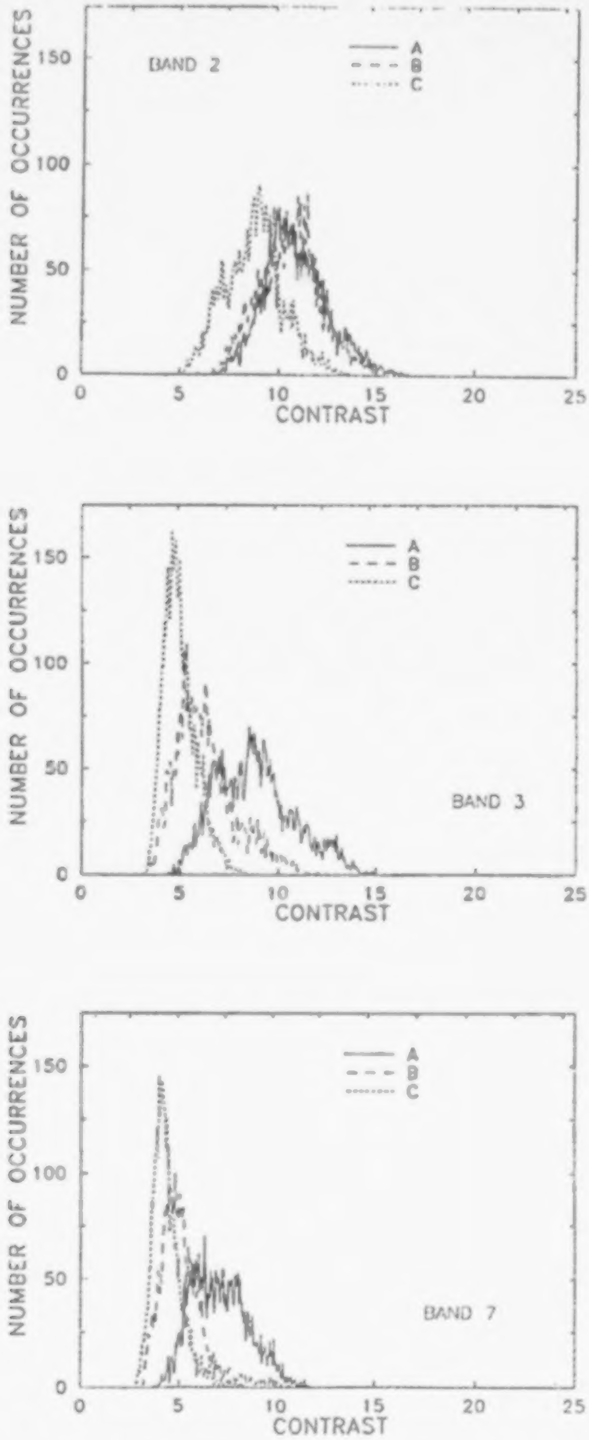


Figure 4: The texture measure Linear Contrast for each of the three bands and each of the regeneration classes.

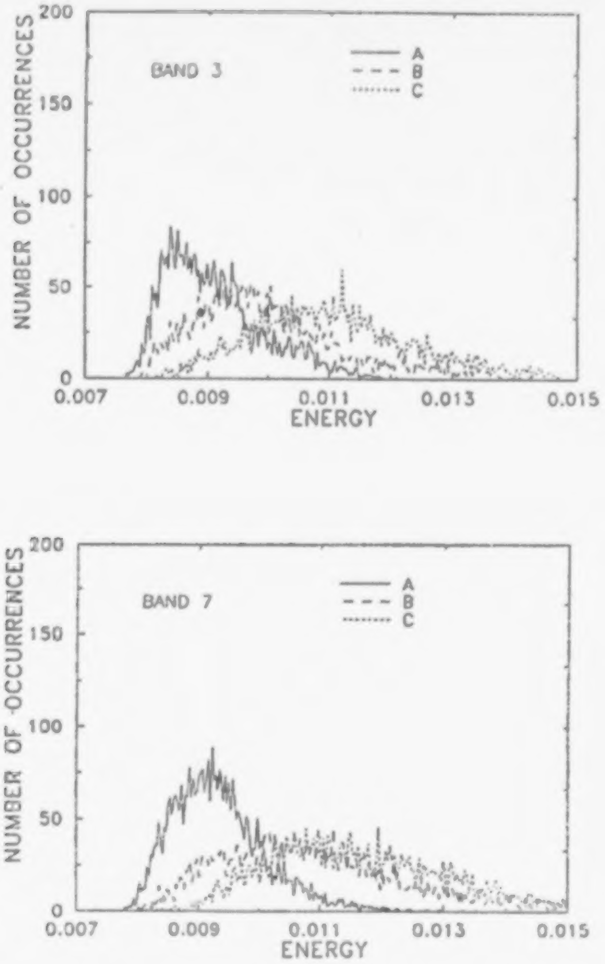


Figure 5: The texture measure Energy for bands 3 and 7 for each of the regeneration classes.

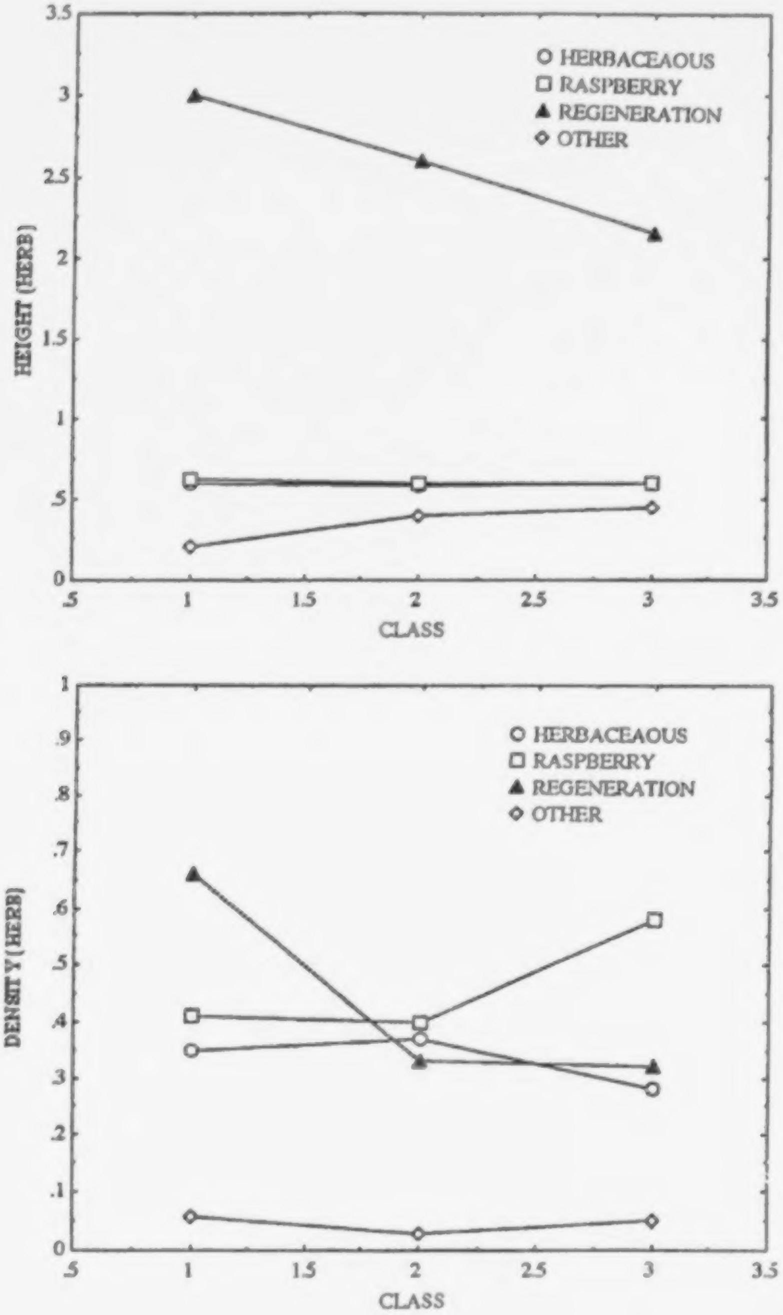


Figure 6: (a) The relative heights of regenerating conifers versus the competing vegetation at the time of the 1989 ground inventory; (b) the relative densities for the same cover types as in (a).

since the others are simply the transposes of these angles and would add nothing to the specification of the texture. Further details can be found in Haralick (1973) and Conners and Harlow (1980). Given the cooccurence matrix, a number of different measurements can be extracted, of which a subset are statistically independent. Five of these measures were studied in this experiment: contrast, energy, entropy, cluster shade and cluster prominence.

Figure 4 shows the histograms of the contrast measurements for each band. These diagrams clearly demonstrate that there is information present in the texture measurements which is not picked up by the purely spectral distributions. This is most clear for band 7, where the intensity distributions for each class were indistinguishable. In the case of the linear contrast measurement for band 7, a clear separability emerges, and one which is stronger than either that provided by the other two bands. Furthermore, for all three bands, the highest contrast is found for the class with the least regeneration. This is intuitively reasonable — we expect the class with the most complete regeneration cover to be the most uniform.

The band 2 results are the worst for all texture measurements. This may be due to the lower contrast between vegetation and shadow in this band compared to both of the other bands. Band 7 appears to give slightly better results than band 3 for most of the texture measurements. For example, Figure 5 shows the energy histograms for bands 3 and 7. On the other hand, as noted above, of the two bands, only band 3 shows spectral differences between the different classes. Hence if spectral measurements are to be combined with texture measurements, band 3 is the best single choice out of the three bands studied for this experiment.

Furthermore, regeneration classes A and C were fully separated, whereas class B demonstrated considerable overlap with each of the two extreme classes. Hence for individual measures (individual windows), only two classes can be clearly distinguished. In order to find intermediate stocking conditions, spatial averaging must take place. This is probably possible and might lead to better results, but is the subject of a later study.

Neither the cluster shade nor the cluster prominence showed significant differences between the classes beyond those already observed in the other measures. These two measures were introduced after it was realized that the earlier measures did not extract the full information from the cooccurence matrix. Empirical verification indicated that these two measures seemed to complete the set of measures (Conners and Harlow, 1980). However, they are much more demanding in CPU time than the other cooccurence measures and hence should be avoided unless they add significant additional information.

Finally, Figure 6a shows the relative heights of the competing vegetation versus the (1989) inventoried heights of the conifer regeneration, whereas Figure 6b shows the relative densities for each cover type. It is clear that the observed correlations between texture and regeneration cover are not due to spurious effects related to the competing vegetation.

It should be noted that the texture analysis was carried out using software developed in-house on an Interpro 340 UNIX workstation using optimization techniques to improve CPU efficiency, since cooccurence measures have a reputation as being very demanding in CPU time.

## Discussion and conclusions

The analysis has clearly shown that textural measurements can be used to measure the percent stocking of regeneration after cutting and hence has the potential of serving as an automated measure for determining when intervention is necessary. This is encouraging because texture measurements are likely to be less sensitive to radiometric differences than purely spectral techniques and hence should apply with little human intervention over a much wider range of conditions. Furthermore, by combining spectral and texture measurements, particularly at 750 nm, the strength of the effect can be further enhanced.

There are two weaknesses to this study. The first is the difference in time between the MEIS overflight and the ground inventory. Efforts were undertaken to remove the influence of this time differential, but inevitably some additional uncertainty was introduced. Nevertheless, the strength of the effect found indicates that it is not spurious. The second weakness is that only a single clearcut was studied. The work *does* need to be generalized to a wider variety of conditions, but the goal of this experiment was to test for the existence of a relationship between texture and regeneration. The results are very encouraging and should be pursued.

## Acknowledgements

We would like to thank the Association des Industries Forestières du Québec and the Natural Science and Engineering Research Council for their financial support for this work within the Industrial Research Chair in Geomatics applied to forestry. We would also like to thank Jacques Brunelle for giving us access to the inventory data he acquired during his M.Sc. degree.

## References

Brunelle, Jacques, 1991, "Contamination de la radiance apparente et extraction de signatures spectrales dans un environment composé d'un système d'information géographique et d'un système d'analyse d'image: cas de la régénération forestière", *Mémoire de maîtrise*, Université Laval, Québec.

Conners, R.W. and C.A. Harlow, 1980, "A Theoretical Comparison of Texture Algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, 204-222.

Edwards, G., R. Landry and K.P.B. Thomson, 1988, "Texture Analysis of Forest Regeneration Sites in High Resolution SAR Imagery", *Proceedings of IGARSS'88 Symposium*, Edinburgh, Scotland, 1355-1360.

Haralick, R.M., K. Shanmugam and I. Dinstein, 1973, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-3, 610-622.

Kneppeck, I.D. and F.J. Ahern, 1987, "Evaluation of a Multispectral Linear Array Sensor for Assessing Juvenile Stand Conditions", *Proceedings of the 20th International Symposium on Remote Sensing of the Environment*, Michigan, 955-969.

Labonté, Martin, Ferdinand Bonn, Gilles-H Lemieux and Gilles Daoust, 1987, "Évaluation préliminaire des données du capteur MEIS II pour la cartographie des surfaces forestières en régénération naturelle après exploitation", Proceedings of the 11th Canadian Symposium of Remote Sensing, Waterloo, Ontario, 239-246.

Till, S.M., W.D. McColl and R.A. Neville, 1983, "Development, Field Performance and Evaluation of the MEIS II Multi-detector Electro-optical Imaging Scanner", *Proceedings of the 17th International Symposium on Remote Sensing of the Environment*, Ann Arbor, Michigan, 1137-1146.

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# An expert system for helping forest managers choose among data sources of remotely sensed imagery

**G. Edwards**

**O. Dhuzel**

**K.P.B. Thomson**

Centre de recherche en géomatique
Faculté de foresterie et géomatique
Pavillon Casault, Université Laval
Ste-Foy, Québec, G1K 7P4
FAX: (418) 656-7411

## Abstract

*Remote sensing is undergoing a rapid expansion in terms of the number and variety of sensors available for information extraction. Furthermore, the increased integration between GIS and image processing technologies is facilitating the use of these images for a wider set of tasks by a larger group of people than has heretofore been possible. However, choosing among the different options has become somewhat of an obstacle course for the non-expert in remote sensing data analysis. This problem is further complicated by the fact that the choice of data is highly task-dependent. A fully functional prototype expert system has been developed for a specific subset of tasks, those related to wood harvesting and regeneration monitoring in the forest management context. The system needed to contain both expertise related to the task domain as well as expertise from the domain of image analysis. The forestry expertise used was specific to the Québec forestry context. The system counsels the user on the "best" single data source or some combination of data sources within constraints such as cost, quality and availability. This prototype was developed in order to explore requirements for structuring the expertise and the relationships between the remote sensing and target domain knowledge for systems of this nature.*

## Introduction

Expert systems have been used on several occasions recently (Schowengerdt and Wang, 1989; Skidmore, 1989) to increase the capability of pixel-based remote sensing image analysis to recognize image objects of interest and to produce map representations of these objects. Furthermore, they are beginning to find a role for interfacing remote sensing image analysis with complex modelling domains such as in global change monitoring. An area which has received less interest, however, is that of knowledge-based systems for aiding the non-expert manager to increase his or her access to remotely sensed imagery and its analysis. This area is very important both for the domestic use of remote sensing imagery and also for international sales. Third world users of remotely sensed imagery must be able to maintain technical expertise over time if they are to profit from the technology. Knowledge-based management systems can help them do so.

Knowledge-based systems may serve in several ways, depending on the context:

(i) as a training system, for teaching new personnel different aspects of the system;

(ii) as a counselling system, for helping knowledge able professionals make informed decisions quickly;

(iii) as a knowledge repository, for storing information about processes and operations which are important.

In this paper, we report on a small prototype system designed to help forest managers select the optimal set of data for different combinations of tasks and also designed to serve as a knowledge repository for the use of remote sensing imagery in forest management.

## Knowledge acquisition and representation

This knowledge-based system was developed in the context of the Québec forest industry, for which information was readily available. In the early stages of knowledge acquisition, a survey of all forest management tasks was carried out. The goal, however, was to determine in which areas image interpretation was most critical, and the later stages were focussed exclusively on detailed knowledge acquisition in these areas.

Since the selection of imagery and data sources is task dependent (as are most image analysis techniques), knowledge of the application domain is *as* important, if not more so, than knowledge of the strengths and weaknesses of different types of imagery and image processing algorithms for those tasks.

It was therefore necessary to characterize the forest domain knowledge *first*, before proceeding to link this knowledge with knowledge about images. For each forest management task, different possible data sources were identified. The data sources were then collected together to produce the inverse table, giving the tasks which are realizable given each data source or combination of data sources.

It was quickly realized that this kind of knowledge-based system is much more a repository of knowledge than it is a sophisticated reasoning system, at least at this level of development. Most of the difficulties encountered were related to the question of how best to *represent* the facts and the relationships between facts, and not so much how to model the intermediate concepts used to derive the conclusions from the input information.

For example, the data sources needed to be rated in terms of the quality of their information as a function of the different management tasks. For example, one task may need a high resolution image in the visible, whereas a different task may only need to find a region of interest and not obtain a detailed map, and hence may be able to benefit from a satellite-based radar image, a cheaper data source.

In general, our experience with other knowledge representation tasks in remote sensing suggest that expert systems developed for remote sensing applications are likely to be wide in the extent of encoded facts, but shallow in the extent of their reasoning or knowledge. Hence efforts directed towards building expert systems for remote sensing applications should concentrate on fact representation issues and the knowledge which represents the linkages between the facts.

## The structure of the system and its reasoning processes

In general terms, an expert system consists of an *inference engine*, a *fact base* and a *knowledge base*. The inference engine can for *forward chaining* (inductive), *backward chaining* (deductive) or *mixed*. There are also at least two kinds of knowledge: well-defined knowledge, including verifiable facts, and intuitive knowledge, which are facts used in intuitive reasoning but not explicitly spelled out by the expert. Sometimes these latter are very difficult to obtain. The most common formalism for representing knowledge is as a set of IF-THEN (so-called) *production rules*. An alternative formalism sometimes used is called *frame-based*, where knowledge is stored in a hierarchical structure of inheritance. These days, most rule-based systems contain some frame-like formalism in order to facilitate knowledge representation. Regardless of the formalism used, it is usually possible to introduce uncertainty into the reasoning process.

Many *expert system shells* exist on the marketplace. These consist of a knowledge representation language or formalism, an inference engine and a variety of editing tools for creating and maintaining knowledge bases.

Also a wide body of literature exists dealing with knowledge representation issues. In simple terms, one needs to model the intrants and extrants of the expert system as well as its different parts and also one needs to model the reasoning dependence chain, including the use of intuitive concepts used to represent intermediate states of the reasoning process. Despite the existence of many tools on the marketplace and a wide body of literature, the development of an expert system still appears a daunting task to the beginner. Although to some extent this is a misconception of the difficulties involved, part of the problem is that the literature is full of warnings of the difficulty of the task and also does not contain much information on simple models to help structure the knowledge base. To a large extent, developing an expert system today is still an exercise in trial and error.

For the system presented here, we used the expert system shell developed and sold by Software Architecture and Engineering, the Knowledge Engineering System (KES). The shell runs on our Interpro340 UNIX workstation. This is a rule-based, backward-chaining shell

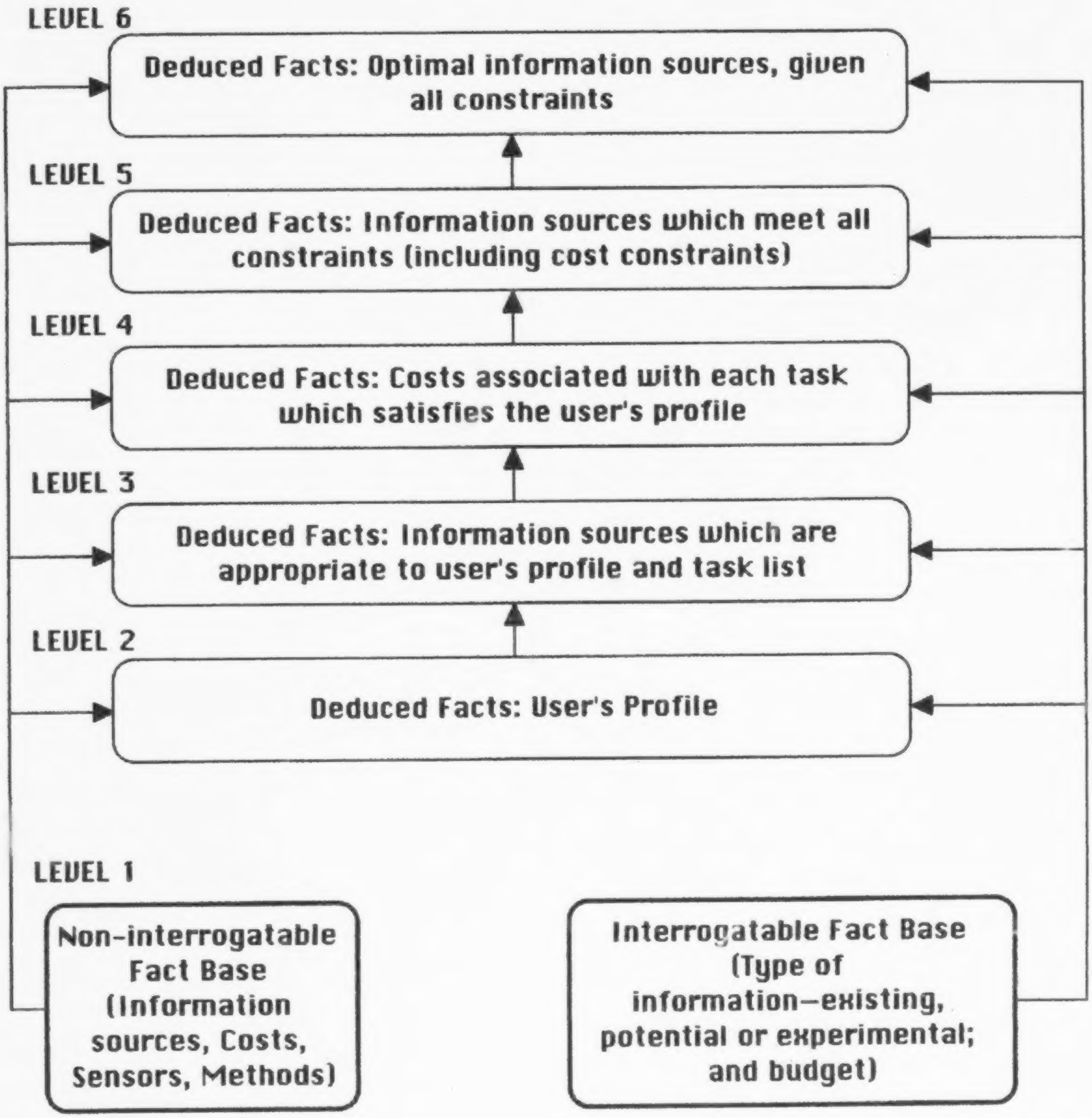

LEVEL 6
Deduced Facts: Optimal information sources, given all constraints
LEVEL 5
Deduced Facts: Information sources which meet all constraints (including cost constraints)
LEVEL 4
Deduced Facts: Costs associated with each task which satisfies the user's profile
LEVEL 3
Deduced Facts: Information sources which are appropriate to user's profile and task list
LEVEL 2
Deduced Facts: User's Profile
LEVEL 1
Non-interrogatable Fact Base (Information sources, Costs, Sensors, Methods)
Interrogatable Fact Base (Type of information—existing, potential or experimental; and budget)

Figure 1: The structure of the expert system. The expert system can be interrogated at any of the six levels, although only at the higher levels does the system perform any significant reasoning. The higher levels use facts derived at the lower levels in their reasoning processes.

with some ability to handle classes (e.g. "frames"). It includes forward-chaining capability via the use of *demons*, as well as the possibility of fuzzy reasoning.

The fact base of forest tasks was built around the norms required by the new Québec forest law. These norms deal with all aspects of forest planning, harvesting and management. During the knowledge acquisition process, however, two major tasks were identified as profiting the most from information derived from images: forest inventory and cut over monitoring (including regeneration). The fact base also includes information on existing image interpretation procedures. Remote sensing image interpretation capabilities were divided into three areas: (i) existing, operational capabilities; (ii) proven but not yet operational capabilities; and (iii) experimental techniques which show promise. The first category includes photo-interpretation of aerial photos and was also widened to include ground-based inventory so as to cover all existing information sources. The expert system was designed to use these categories as constraints. A "user-profile" is constructed by the system after querying the user, in order to determine if he or she wishes to confine the system to existing practice or if he or she is interested in using new but not yet operational techniques, or would like to see what still experimental techniques show promise.

The elaboration of the logical dependencies between the information sources and the forest tasks led to a massive data base containing more than 40 pages of coded relationships. The types of information included include maximum and minimum image resolution for each defined task, the types of image processing or enhancement needed to extract the information, the types of sensor appropriate to the tasks, and so forth.

The next level of dependance consisted of determining the relative costs of different data sources, given the information acquisition requirements in terms of temporal window, time of year, mean local weather conditions, nature of the platform (airborne or spaceborne), etc.

The major part of the "reasoning" used by our system consists of the rule-base for determining an "optimal" set of images for a given set of tasks, based on heuristics similar to what a human uses, and given the different knowledge constraints stored in the system. Hence the system determines the availability of the different sensors, based on global weather information for the given area and evaluates the availability as a function of the window constraints of each forest management task for which the information is required and in terms of an overall budget. The system then suggests a prioritized series of acceptable images. The constraints can be modified at will to produce new suggestions.

Demons, which are triggered as soon as the antecedant condition is filled, were used almost exclusively to guide the user in the choice of tasks to be carried out. The expert system permits the user to begin with general task categories and to progressively refine the requirements to obtain specific tasks.

The overall structure of the system is shown in Figure 1. This structure is of interest because not only does it provide a solid fact base from which reasoning can be launched but it can be interrogated at any level, hence serving as a flexible knowledge repository. Because the modular design allows reasoning to be broken down into layers, the system is, in principle, more maintainable and more easily extended than is usually the case with expert systems. An example of the "flow" of a typical session is shown in Figure 2.

## Discussion and conclusions

The complexity of modern image analysis tasks resides largely in the wide variety of options and choices available to the user. Hence the major difficulty for the non-expert is to choose the appropriate procedure for a given task or set of tasks. An expert system developed to help the non-expert choose must initially cover a large variety of situations, rather than follow a long chain of deductions. The early development should therefore focus on the relationships between facts and the representation of these relationships, especially if the user is to have access to this fact base. Once a solid base has been established, it is relatively easy to extend the depth of the reasoning to cover additional tasks, such as obtaining solutions under a variety of different constraints. In this work, a modular level approach to representing reasoning was successfully used. Such an approach is especially important considering the size of the fact base. Without such structure, the knowledge representation could get very complicated and hence difficult to either verify or maintain.

Expert systems for advising the non-expert in image processing and/or GIS use are almost certainly of great practical value, but they do not extend the capability of an expert to use these technologies. Rather, they help standardize procedures and help train new "experts" by giving them informed decisions on the choice of images and appropriate image processing or enhancement techniques. They also provide a valuable repository of knowledge. As such, the development of these systems is likely to be of special interest to private industry.

## Acknowledgments

We would like to thank the Association des industries forestières du Québec (AIFQ) and the Natural Science and Engineering Research Council of Canada (NSERC) for their financial support of this work within the framework of the Industrial Chair in Geomatics Applied to Forestry.

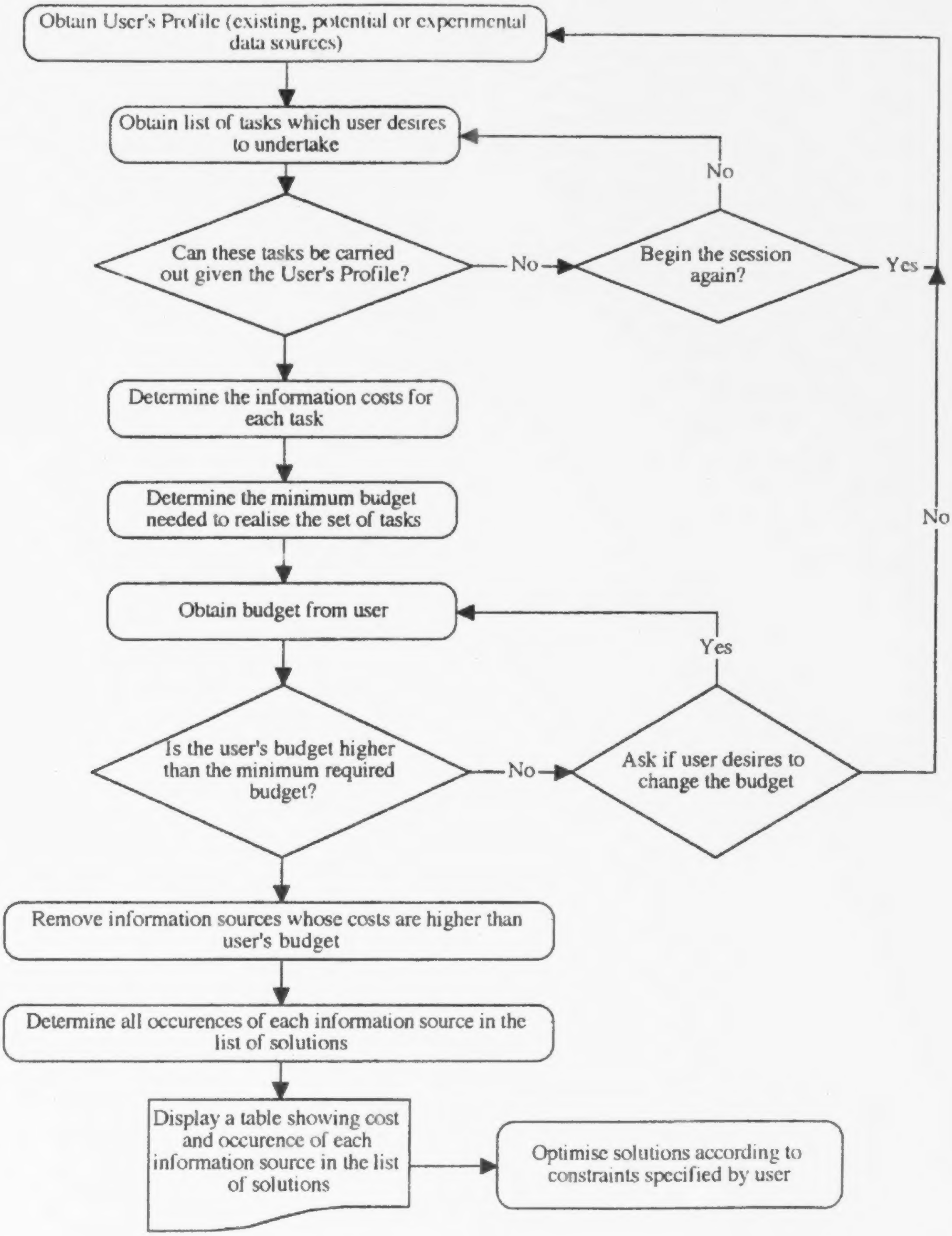

Obtain User's Profile (existing, potential or experimental data sources)
Obtain list of tasks which user desires to undertake
No
Can these tasks be carried out given the User's Profile?
No
Begin the session again?
Yes
Determine the information costs for each task
Determine the minimum budget needed to realise the set of tasks
No
Obtain budget from user
Yes
Is the user's budget higher than the minimum required budget?
No
Ask if user desires to change the budget
Remove information sources whose costs are higher than user's budget
Determine all occurences of each information source in the list of solutions
Display a table showing cost and occurence of each information source in the list of solutions
Optimise solutions according to constraints specified by user

Figure 2: Chart showing the flow of a session with the expert system. The last box contains most of the reasoning in the system. The remainder of the system is mostly a knowledge repository.

## References

Schowengerdt, Robert A. and Han-lung Wang, 1989, "A General Purpose Expert System for Image Processing", *Photogrammetric Engineering and Remote Sensing*, Vol. 55, 1277-1284.

Skidmore, Andrew K., 1989, "An Expert System Classifies Eucalypt Forest Types Using Thematic Mapper Data and a Digital Terrain Model", *Photogrammetric Engineering and Remote Sensing*, Vol. 55, 1449-1464.

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# A geographic information system for strategic and operational monitoring of forestry activity: a conceptual approach

**Edi Sutopo**
**Riadika Mastra**

PT Mapindo Parama
JL. Bendungan Hilir Raya No. 50
Jakarta 10210
Indonesia

## Abstract

*In the managerial process, the main differences between strategic issues and operational ones are mainly: the finishing time, the degree of detail needed in data and the insight of the information. The strategic process needs little time to get adequate information so that good decisions can be made. For the operational process the differences are in data volume, data organization and the method of analysis.*

*Therefore data must be identified as strategic data and operational data in the same GIS. Nevertheless, both still have a connection to each other when updating data: for example, whenever operational data is updated, strategic data must be updated automatically. Updating operational data into strategic data is done by the Generalization process.*

*Both data have different characteristics, to the extent that operational data has detail and strategic data has a global perspective. This makes a big difference to the volume of the two kinds of data; which results, for example, in different scale maps, different kinds of data displayed, the grade of accuracy of the data, etc. Therefore, one must design two different kinds of organizations for the data in a single GIS system.*

## Introduction

The area of forest in Indonesia is 143 million hectares which is about 74% of the total land area in Indonesia. There four kinds of forest in Indonesia (see next page).

The *production forest* is managed using the HPH (Hak Pengusahaan Hutan = the forest concessionaires) system. At the moment there are 574 HPH companies, 524 of which are united in an Association called APHI (Asosiasi Pengusaha Hutan Indonesia = The Association of Forest Concessionaires of Indonesia). The Government of Indonesia conducts forest inventory, especially product forest inventory, through this APHI. Later on the government, through APHI, may also conduct inventory of other kinds of forest.

The government has decided to use a forest management system which is based on a work plan. This work plan consists of a long-term work plan called RKL (Rencana Karya Lima Tahun = A Five Year Work Plan) and a short-term

| | |
|---|---|
| Protected forests | 30 million hectares |
| Nature reserve and recreation forests | 19 million hectares |
| Production forests | 64 million hectares |
| Converted forests | 30 million hectares |
| | Total 143 million hectares |

work plan called RKT (Rencana Karya Tahunan = An Annual Work Plan). The forest exploitation system applied by HPH is called TPTI (Tebang Pilih Tanam Indonesia = Selective felling and Planting System in Indonesia).

The purpose of the forest inventory in Indonesia is to establish a database of the potential of the forests and thus to obtain information on the area of virgin forest, the area of exploited forest, and the area of critical forest in order to determine appropriate policies and steps concerning the development and the use of forest as a natural resource.

## Forest inventory in Indonesia

Forest inventory in Indonesia is meant to obtain accurate and up to date data on the conditions and potential of the forest. The data are illustrated in form line maps and vegetation maps at a scale 1:25 000.

*Form line maps* are similar to topography maps;only the height lines are the heights above the trees. But on open areas, the height lines are the same as those in topography maps, i.e. the height on the ground.

*Vegetation maps* are maps which provide information on forest condition in the form of forest class map stratification which are illustrated in crown density, tree heights and crown diameter. Each class is further divided into three classes. So the maps produced later on will have 27 data combinations (c1, h1, d1) —(c3, h3, d3). Besides these data, types of vegetation will also be provided.

Data collection for the forest inventory uses panchromatic aerial photos at a scale 1:20 000. The use of panchromatic aerial photos also makes possible photogrammetric mapping and vegetation interpretation. In addition, field research on the aerial photo interpretation is also done.

Aerial photography is applied to all the forest areas which are managed by HPH companies, so that with a map at scale 1:25 000, detailed information on the object in the field can be obtained and good logging can be well prepared.

Forest inventory and analysis also use other supporting data such as:

- Geological map
- Geomorphological map
- Indonesian Forest Land Use (TGHK) map
- Soil map
- Rainfall map
- Land suitability map
- Hydrological map
- Government administration map
- Forest region administration map
- Forest exploitation work plan map
- Cutting location map
- Replanting location map
- etc.

## Monitoring forest exploitation in Indonesia

The form line maps and vegetation maps provide us with detailed and most up to date information on the condition of forests in a certain area, in this case, the forest which is managed by a HPH company.

As has been discussed previously, forest exploitation plans in Indonesia are realized in the forms of RKL and RKT which are planned on those maps above. The exploitation plan was proposed by HPH in the form of RKPH maps (Rancana Karya Pengusahaan Hutan = Forest Management Work Plan) which was further proposed to the Directorate of Forest Management at the Department of Forestry.

The basic concept of forest exploitation is maximum usage of forest products without damage to the environment, which means the environment should be well protected. Therefore, a system will be established so that forest exploitation activities in Indonesia can be closely monitored.

To determine the degree of development of the forest, and the result of exploitation of the forest managed by HPH companies, a monitoring system is necessary. The monitoring system is divided into two parts. Firstly, the system to be applied for data collection for monitoring purposes, e.g. by taking aerial photos every five years, or by using GPS for positioning any change taking place, is based on the annual work plan or by using Landst TM, SPOT Image, Radar Image. Secondly, how to organize the data of those changes towards the existing data. A Geographic Information System (GIS) provides a very good system for monitoring purposes as it can include spatial as well as textual information.

Moreover, because of the use of TPTI system for forest exploitation, a system will also be established to monitor the application of the TPTI or to find out whether the TPTI has been done correctly or not.

Therefore, it is necessary to develop a Geographic Information System (GIS) which is suitable for forest analysis purposes. If the GIS is developed and managed by APHI, APHI will be able to monitor forest exploitation done by either its members or other HPH companies.

## GIS for APHI

GIS managed by APHI will be formed in two different integrated models. The first is an operational mode which concerns problems directly related to HPH. Secondly, APHI is supposed to be able to provide strategic information needed for decision-making by leaders of government bodies - such as the Department of Forestry, which is concerned with forest potential and all similar aspects; the Department of Trade, which is concerned with export commodity potential, BAPPENAS (The Department of National Planning) which is concerned with problems of national and regional development, the Department of Industry, which deals with development of industry from the forest sector. This kind of model is called a strategic model.

### Strategic an operational model GIS

The forest inventory will cover a forest of 64 000 000 hectares throughout Indonesia. As has been indicated previously, the basic map is made at a scale of 1:25 000 which includes 574 HPH companies or 1128 locations. With such a big volume of data, a filing system for the back-up should be considered.

APHI as an organization which organizes the HPH companies will provide the government with inputs related to forestry and timber in Indonesia in its effort to protect the environment and maintain the forest as a renewable natural resource since forest and wood products are the main sources of national income. So APHI must be able to provide accurate, clean and comprehensible information whenever needed.

For these reasons, an integrated and distributed GIS should be established. An integrated GIS means that the GIS is managed by APHI. It provides information for the government through its departments and directorates. The integrated GIS also makes it easy for the APHI to manage its members. A distributed GIS is managed by an APHI member for the importance of its own HPH company. Each distributed GIS must have data with a standard form because the data is used as updating data in APHI. The Integrated GIS managed by APHI has two kinds of data which are different but use the same system. They are *strategic level data* and *operational level data.* These two levels of information are very different in their scales as well as their characteristics.

Strategic level data is derived from operational data in a certain way called *Generalization.* During the generalization process there are changes in the scale and the characteristics. The Generalization is divided into two: Spatial Data and Textual Data Generalizations.

## Generalization

Spatial Generalization is caused by change in scale. For example, in operational data at a bigger scale there are several classes of road: country road, province road, regency road, and other lower class roads. As the Spatial Generalization is caused by the change in scale, from a bigger to smaller scale, some lowest class roads are omitted, so there are only the main roads in the supporting information strategic data. It is the same case with data on rivers, villages etc.

Other common Generalization forms such as line coordinate thinning, polygon coordinate thinning, drop line thinning and edge matching are also used in Generalization because the main principle of Generalization is reduction of data volume. A smaller data volume will produce faster data access and thus immediate strategic information can be presented.

Textual Generalization is a Generalization with statistic process or summaries. See example below.

Textual Generalization causes changes in data grouping limits like in reclassification in abstraction map.

The Generalization process may cause a decrease in the quality of data but not in the accuracy.

| Group (year) | Age of tree | | Group | Age of tree (year) |
|---|---|---|---|---|
| 1 | 0 - 5 | Generalization | | |
| 2 | 5 - 10 | ——————> | 1 | 0 - 10 |
| 3 | 10 - 15 | | 2 | > 10 |
| 4 | > 15 | | | |

## Application

In integrated strategic and operational GIS, several applications which provide information can be identified.

**Strategic applications**

- History of forest development

  This keeps data on forest annual development which will help in the animation of forest annual development.
- Forest location map
- Virgin Forest Area
- Potentialities non virgin forest
- No potentiality non virgin forest
- After cutting
- Critical forest map
- Forest potentiality
- Throughout Indonesia
- Commercial timber
- Non commercial timber
- Industrial timber estate
- Forest production potentiality
- Annually
- A five year period
- Anually per timber type
- A five year period per timber type
- Export potentialities
- Annually
- A five year period
- Natural reserve distribution map
- Conversion forest realization
- Production potentiality of timber industry
- Raw material resource for timber industry
- Production capacity of timber industry
- Potentiality of industrial timber estate
- Indonesian replanting growth throughout Indonesia

**Operational applications**

- RKL and RKT Planning and realization
- HPH working map updating and editing
- The forest policy changes realization

## Distributed GIS application

**Strategic**

- Long term timber stock
- Long term cutting plan
- Annual timber cutting budget plan annually
- etc.

**Operational**

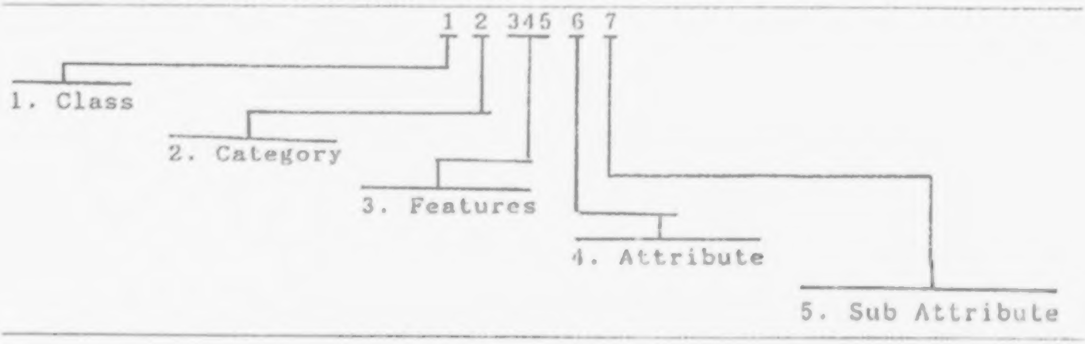
- cutting block lay out plan
- Logging road plan
- Seedbed implementation
- replanting
- etc.

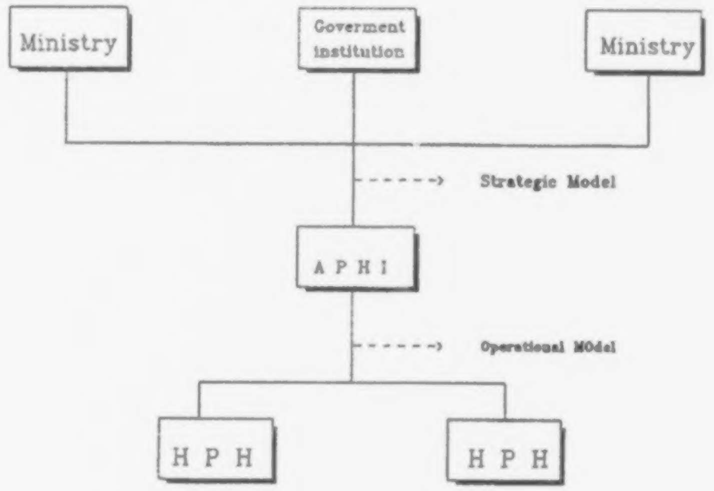
## Data base concept

Before both the strategic and operational subsystem is operated, the data base for supporting both sub-systems should be established. This data base concept it is very important; if the the data which is already stored is to be called for editing purposes, all data must be in organized forms. To do so, organization of the input data should be carried out beforehand. In order to organize the input data, classification and coding of the input data should be done in the beginning. For this purpose, the relationship between input data and the data base itself should be cleared first. Figure 7. shows the relationship between the model of the geographical data and the data base.

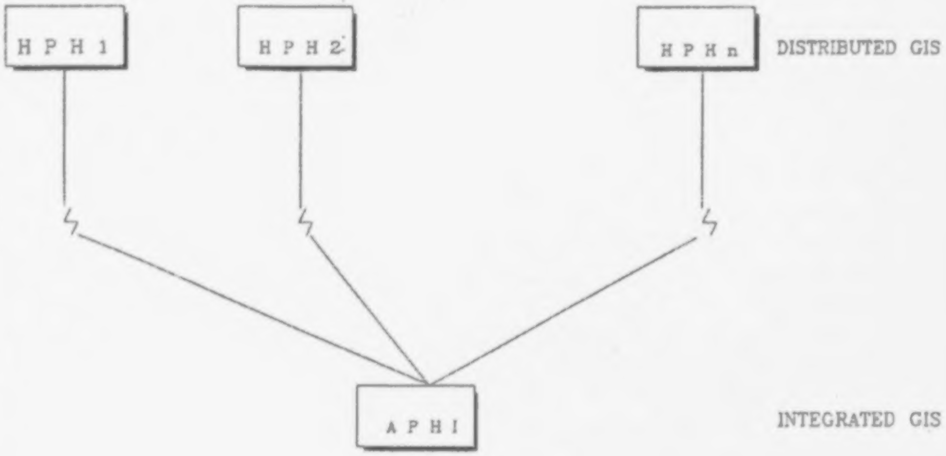
From figure 7, group data can be established as follows:

1. Forest Administration data
2. Indonesian Forest Land Use (TGHK)
3. Forest Concessionaire data
4. Forest Potential data

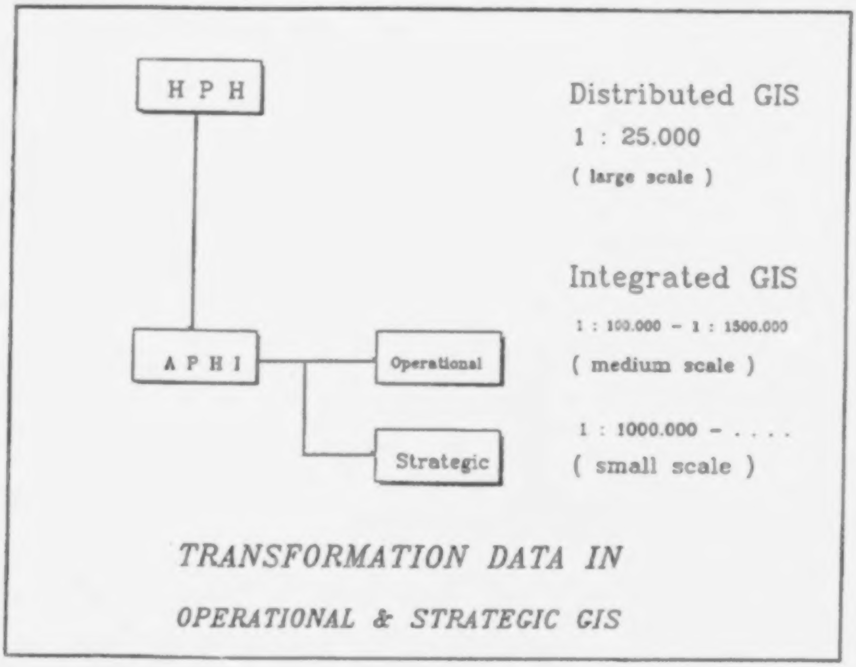



Ministry
Goverment institution
Ministry
Strategic Model
A P H I
Operational MOdel
H P H
H P H

*STRATEGIC & OPERATIONAL GIS MODEL*

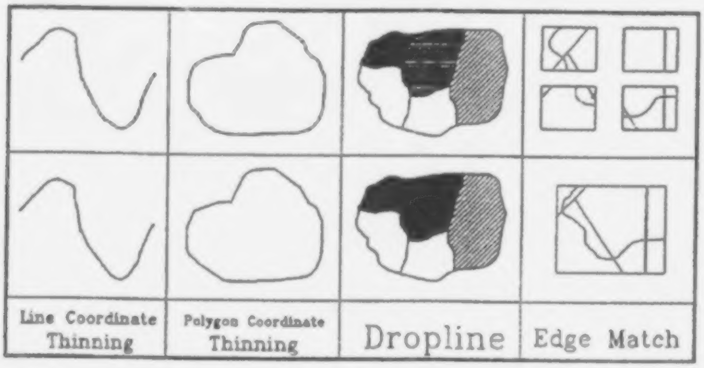

H P H 1
H P H 2
H P H n
DISTRIBUTED GIS
A P H I
INTEGRATED GIS

*INTEGRATED & DISTRIBUTED GIS*

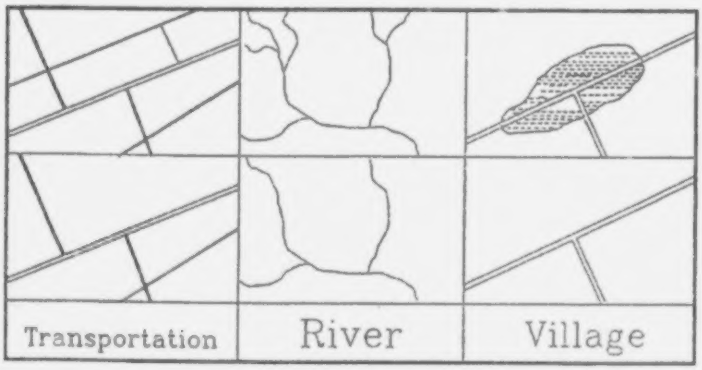

H P H
A P H I
Operational
Strategic
Distributed GIS
1 : 25.000
( large scale )
Integrated GIS
1 : 100.000 – 1 : 1500.000
( medium scale )
1 : 1000.000 – . . . .
( small scale )

*TRANSFORMATION DATA IN*

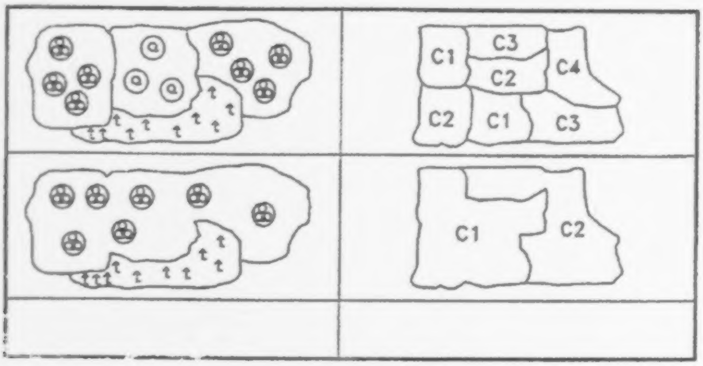
*OPERATIONAL & STRATEGIC GIS*


Line Coordinate Thinning
Polygon Coordinate Thinning
Dropline
Edge Match

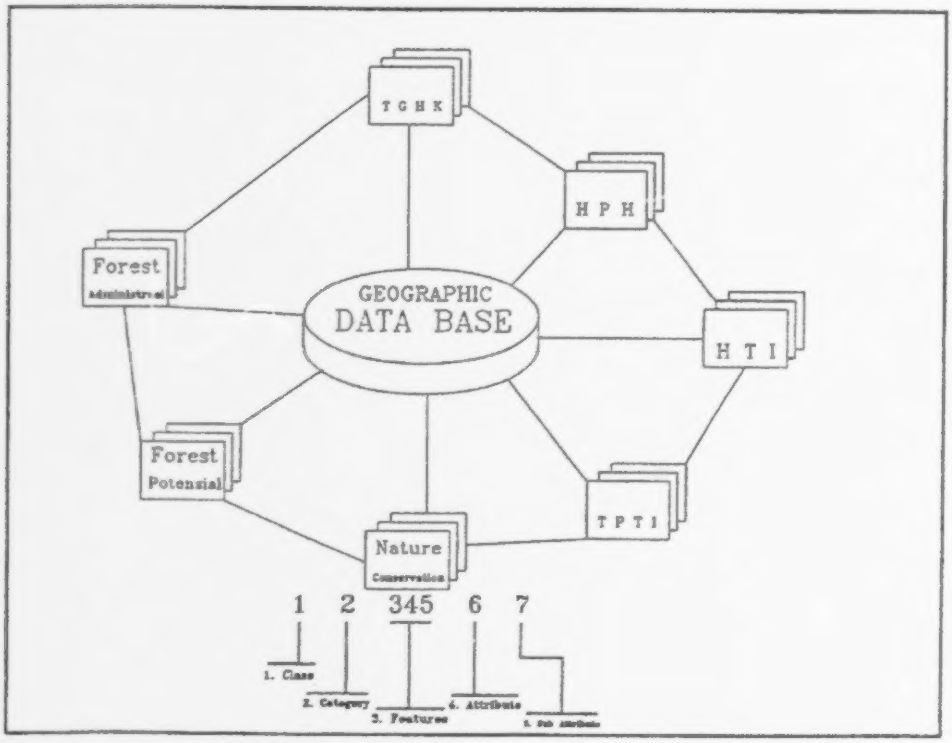
*MAP GENERALIZATION*


Transportation
River
Village

*SPATIAL GENERALIZATION*


t
C1
C3
C2
C4
C2
C1
C3
C1
C2

TEXTUAL GENERALIZATION


T G H K
H P H
Forest
Administrasi
GEOGRAPHIC
DATA BASE
H T I
Forest
Potensial
T P T I
Nature
Conservation
1
2
345
6
7
1. Class
2. Category
3. Features
4. Attribute
5. Sub Attribute

| No. | Class<br>(A-Z) | Category<br>(A-Z) | Features<br>(000-999) | Attribute<br>(A-Z) | Sub Attribute<br>(A-Z) | Code<br>1 2 345 6 7 |
|---|---|---|---|---|---|---|
| 1. | South Kalimantan | Forest Potential | Red meranti | Commercial | 20-30 m | A E 146 A B |
| 2. | | | | | | |

5. Industrial Timber Estate
6. Natural Conservation data
7. Selective Felling and Planting System in Indonesia

Each of the group data consists of many layers of data and lists of informations. To organize these data, codefication of each item of information is necessary. In order to give a code of each feature from the group data, each data should be put in a certain group; 5 (five) groups of data are established.

1. Class
2. Category
3. Features
4. Attribute
5. Sub-Attribute

During inputting of data into the database, the structure code of the five group of data is shown below.

The Class consists of 26 items (A-Z)

The Category consists of 26 items (A-Z)

The Features consists of 999 items (000-999)

The Attribute consists of 26 items (A-Z)

The Sub-Attribute consists of 26 items (A-Z)

For example, if the input data consist of:

1. South Kalimanatan province (A-Z)
2. Has data form Forest Potential (A-Z)
3. Information about type of tree (red meranti = Shorea acuminatadyer from Dipterocarpaceae) (000-999)
4. Has a commercial value (A-Z)
5. With height information of the tree (20-30m) (A-Z)

This can be coded as A E 146 A B, the information is taken from South Kalimantan province, from category forest potential data, according to the established type of commercial tree is number 146, has a commercial value and during the investigation time the tree has height about 20 -30 m height.

All of this kind of information is listed and distributed to the operational subsystem section in order to maintain the uniforminity of the input data. The list supporting this information can be illustrated as shown above.

The class itself consists of all province names which have relationships with forest concession activity, and for Indonesia it consists of 27 provinces, the category consists of seven from figure 7. and the addition of topography, climatogy, geology and soil type.

## Summary

The Indonesian forest is a tropical rain forest which has various species of trees and animals. The forest is a renewable natural resource. Such a natural environment must be protected. Therefore, forest exploitation must always be followed by recovery effort. The Indonesian government has determined policies concerning forest management through HPH companies with selective felling and planting system (TPTI). The TPTI system should be followed with a monitoring system to find out how far and how well the companies follow the system. This makes APHI (Association of Indonesian Firest Concessionaires), as the government's partner in forest management, propose the development of a forestry GIS.

# Eco-camp an environmental education experience

**Instructors Professor Mark von Wodtke,**
**Associate Professor Timothy R. Day** (author)
**Graduate Students David W. Briley, Charles LaRue,**
**Anne Pietro, Lisa Skillett**

Department of Landscape Architecture,
California State Polytechnic University,
Pomona

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

## Abstract

*In the western United States, where a significant amount of land is owned and managed by the federal government, substantial pressure exits to convert large quantities of this land to urban uses. The pressure on this land is most acutely felt by the most sensitive ecosystems, due to high attractiveness for human uses. Before these ecosystems are irreparably damaged, more attention needs to be focused in the direction of onsite public education in environmental systems. This study proposes the creation of an Eco-camp to provide this learning experience.*

*Since the passage of the Multiple Use Act in 1960, the interaction between National Forest managers and users has been controversial. Short-term environmental awareness programs developed by the U.S. Forest Service and other federal land managers have largely failed to adequately address this need. This paper describes a project in which a Geographic Information System was used to model and determine the location of possible sites for a permanent and sustainable environmental education center in the San Bernardino National Forest, California. While the primary purpose of the center is education, the context of its operation is intended to be recreational in nature.*

## Introduction

Human population growth and migration has resulted in significant changes in environmental impacts and awareness. Pressure to develop and inhabit sensitive land has negatively affected not only our planetary ecosystems but also our land ethic which is jeopardizing our long-term quality of life.

In order to adequately address this problem, there is a need to provide places on our planet where our species can be exposed to the dynamic ecosystematic paradigm. Through continued exposure, humans can acquire not only a better appreciation of our local and global ecosystem but also the knowledge and motivation to embrace and solve a multitude of global environmental problems such as air pollution and loss of topsoil.

## Goal

The purpose of this study is to determine locations that fulfill the criteria for an Eco-camp within a portion of the San Bernardino Mountains in S. California.

Camping is a favorite recreational activity for many cultures. Interest in camping exploded after World War II due in part to the population growth phenomena known as the "baby boom". This has lead to a tremendous increase in the recrational use of our natural environments.

A popular approach to camping is the idea of thematic camping where specific types of activities take place in a setting far away from urban or suburban life-styles. Eco-camping is one type of thematic camping in which the campers participate in a program of activities concerning the ecologies of a visited landscape. Programs vary in camps such as this, but participants report a significantly higher understanding of the planet on which they live and a greater appreciation of the ecological problems created and faced by man. The idea of an Eco-camp in the San Bernardino National Forest is fairly new. There are several 'nature trails' and indoor awareness activities related to forest ecology and use, but none that combine both camping and environmental education.

In addition to camping and environmental education, the locational criteria developed in this study includes the concept of sustainability. Sustainability is not a new concept, especially to foresters. The principles applied here, however, relate to the work of Howard T. Odum (1983) and John T. Lyle (1985) that deal with the energy and material flows that occur in natural systems. Based on these principles, this camp will be designed to promote sustainability and minimize negative impacts from camp facilities and activities. It will also introduce participants to the ecologies of the region and to experience many environmental issues first-hand.

Consequently, while the focus of this camp is educational, it is intended to be experienced by people in a recreational context. The camp concept would include a balance of topical instruction and fun. The assumed audience for the camp would include people from a variety of ages, cultures and socioeconomic sectors.

## The site

### Morphology

The San Bernardino Mountains along with their westerly neighbors, the San Gabriel Mountains, are part of what is known as a transverse range province. The term "transverse" relates to the west to east trending which is counter to the northwest to southeast structural grain of the Sierra Nevada mountain range (Robinson 1979).

Formed by intensive folding and faulting, Clark (1978-1979) describes the San Bernardino Mountains in morphological terms as a precambrian igneous-metamorphic complex outcrop also containing shists, gneiss and migmatites. The mountains are overlain by weakly metamorphosed upper paleozic marine quartzites and limestones. They create a climatic barrier between the mediterranean San Bernardino Valley to the south and the arid Mojave Desert to the north.

### Plant community

Plants are generally distributed based on a north and south aspect, high and low elevations, and next to stream channels. The south facing slopes are generally composed of chaparral including species such as chamise, scrub oak, manzanita, California holly and ceanothus. The rolling crest contains pines (Jeffrey, Ponderosa, sugar, Coulter and Knobcone), incense cedar, white fir and dogwood.

On the desert facing north slopes, the pinyon-juniper woodland is found at the higher elevations . The lower elevations, or lower sonoran desert, contain the Joshua Tree, mountain mahogany, mesquite and sages.

Tree canopy for the stream-side woodland contains sycamore, white alder, willow, big-leaf maple, and live oak while the understory is composed of ferns, rushes and herbs (Robinson 1979).

### Wildlife

Much of the local fauna has either been gradually eliminated by man or has moved to new habitat. The native grizzly left about 50 years ago but a few black bear still remain. A few bighorn sheep still climb the desert south slopes and mountain lions can be seen from time to time. Mule deer are perhaps the most abundant large animal still thriving in the mountains.

Small animals prevalent in the mountains include bobcat, ring-tailed cat, gray fox, coyote, opossum, raccoon, skunk, weasel, squirrels, and chipmunks. Rattlesnakes have been occasionally sighted below 6000 feet (Robinson 1979).

## Camping history of the San Bernardino Mountains

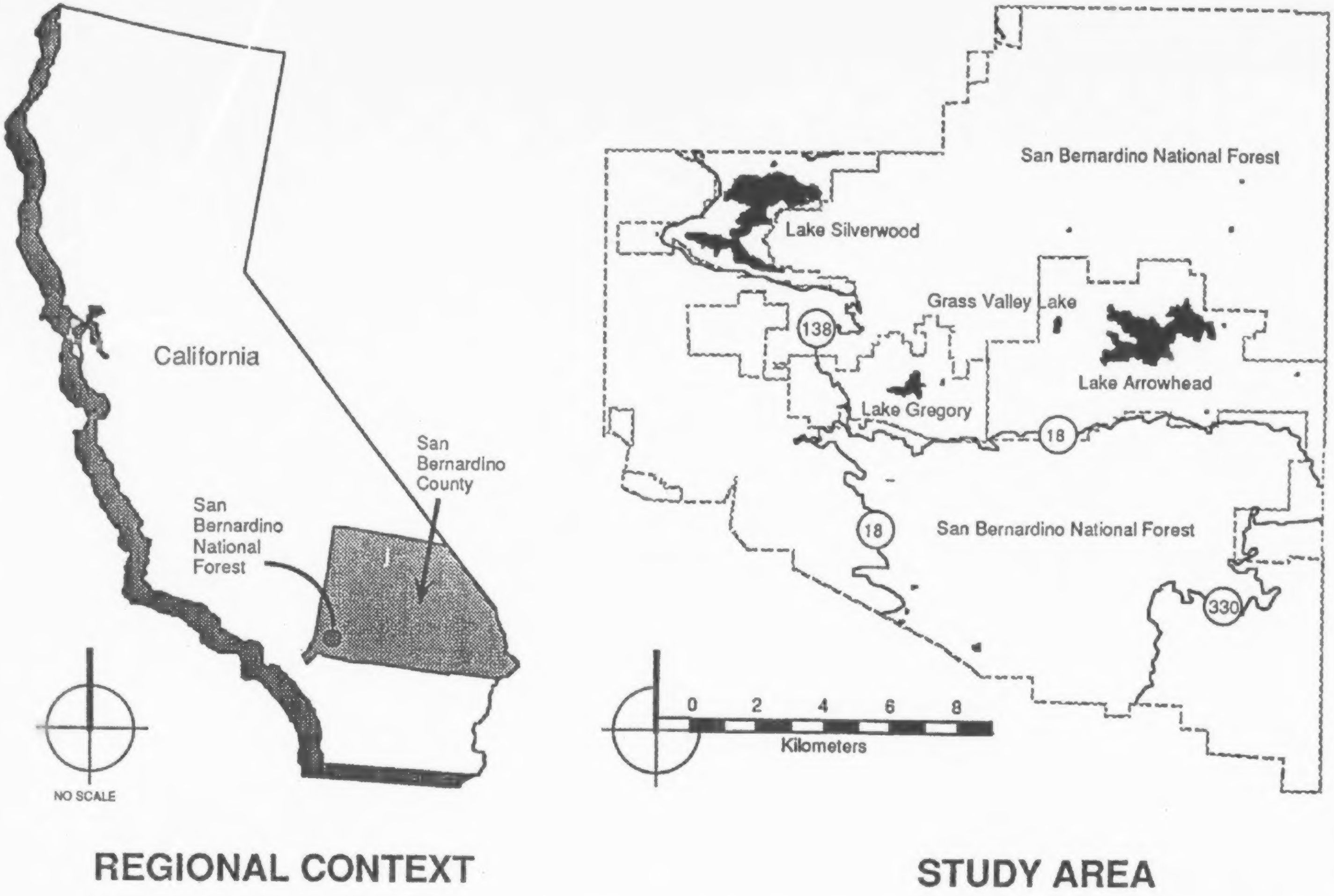
### Primitives

The Indian people that inhabited the mountains were nomadic hunters and gatherers. They belonged to the Shoshonean linguistic category, and were called Serrano which is Spanish for "Mountain People". They dwelt on the southern and northern slopes, alongside streams, around springs and lakes or at the mouths of canyons. The settlements of serranos were called "Rancherias" which means village sites. Robinson (1989). Most settlements containing 10 to 30 dwellings were in the lowlands, with seasonal sites in the high country. One known example of a seasonal site is Rock Camp, which has been a favorite archaeological area for many years. LaFuze (1971) states that they were primarily vegetarian, eating manzanita berries, pinyon pine nuts and acorns ground in "metates" (bowls) in the exposed rock.

### 1800-1900

Glimpses of what life was like on the mountains in the 1870's can be found in the diaries kept by a lumberman by the name of Joseph B. Tyler and reproduced in LaFuze (1971). Recreational activities in those days

**Figure 1:**


California
San Bernardino County
San Bernardino National Forest
NO SCALE
REGIONAL CONTEXT
San Bernardino National Forest
Lake Silverwood
Grass Valley Lake
Lake Arrowhead
Lake Gregory
138
18
18
330
San Bernardino National Forest
0
2
4
6
8
Kilometers
STUDY AREA

were simple. On Sundays, traditionally regarded as a day of rest, Joe either walked to a neighbor for a visit and dinner or just stayed "home all day" reading the news or a book. July 4th was usually celebrated with a picnic and as more people moved into the mountains, a dance at one of the mill cookhouses. In winter, walking became snow-shoeing and we find the quote "had lots of falls and fun." Sewing bees and quilt making were popular. Occasionally residents would hunt deer and other large game.

By the 1880's, there were several primitive mountain camps along the crest of the San Bernardino Mountains. Hikers and campers found their way to places such as Devil Canyon, Seeley Flat or Little Bear. Summer vacations at these rustic camps included boating, fishing, burro riding and even sometimes snow activities at the higher elevations.

Mill Creek Canyon was a favorite camping spot for many fishermen along the Santa Ana River. Robinson (1976) describes the canyon as a "fine mountain torrent wide and bouldery through a charming woodland of Coulter and Yellow pines, incense cedar, white fir, Kellogg oak and maple." There were "Goldenrod, and aster-like stars of corethrogyne, scarlet wild fuchsia, yellow Helenium, Vermilion paintbrush, scarlet penstemon, cardinal Mimulus and Chinquapin on the white soil of the slopes." Fishing in the canyon was so good that in a short time the creek became over-fished. According to LaFuze (1971), the California Fish and Game Authority placed the limit of a day's trout catch in mountain streams at 50, but the sale of hides from local mountain fauna was forbidden.

On February 25, 1893, President Benjamin Harrison signed into law a bill passed by Congress to establish 737,280 acres of public forest reserve in San Bernardino County called the San Bernardino Forest Reserve. Mr. C. S. Newhall was appointed forest guard in 1897.

Several camping areas and resorts were popular including, Seven Oaks, Camp Indolence, Bart Smithson's strawberry farm, Palmer's apple orchard. Mills were also popular including Guernsey's, Talmadge's and Fleming's. It is said that wagonloads of vacationers and campers were not only lining Mill Creek with camps, but also the ferny grottos of Smith Canyon, the cold splashing Cable Creek, and Holcomb Valley.

**1900-1940**

Several camps and lodges were built in Mill Creek Canyon including Camp Kil Kare near the confluence of Mill Creek and Mountain Home Creek. Cedar Lodge and the Rainbow Angling Club were also started. LaFuze (1971) states that fishing was so good in 1903 and 1904 that streams were becoming fouled from the negative impacts of the camps adjacent to them. Sylvanus Thurman was leading climbing parties up "Old Grayback" now known as San Gorgonio Mountain.

Religious services were offered at camps in the Santa Ana, Mill Creek and Coldwater canyons, and a theater was built by Mrs. Mecham and Miss Preciado for plays, dancing, singing and music.

In 1908, President Roosevelt created the Angeles National Forest including the San Bernardino Mountains. Due to the increased popularity of hunting and fishing, the Fish and Game Association was busy stocking streams every Fall and importing grouse and wild turkeys from Mexico and Hungary. The Los Angeles Recreation Department leased most of Seeley Flat from the Forest Service and built Camp Seeley with a lodge, a dining room, cabins and tents, a swimming pool and tennis and croquet courts. Camping was thriving so well that Bill Codd peddled apples, peaches, berries and watermelons to all the camps.

Despite the fact that war was declared on Germany in 1917, recreation in the mountains was busier than ever. During the war, it was said that "labor hours were long; checks were big, and the working people took their vacations like "rich folks" (LaFuze 1971). In 1917, the Pacific Electric built a vacation camp for its employees next to Little Bear Creek. The City of Los Angeles leased a portion of Converse Flats in 1917 to build a recreation center. It opened in 1918 and later became known as Camp Radford.

In the 1920's, the mountains realized the greatest surge of growth in recreation in its history. In May 1925, the San Bernardino National Forest was divided from the Angeles National Forest by President Coolidge. Steven Augustus Nash Boulden became the first supervisor of the newly created San Bernardino National Forest. He issued the first Forest Recreation Plan to make the mountains "one of the foremost playgrounds in the nation" (LaFuze 1971). Supervisor Boulden announced a 500% increase in campground and summer home facilities including a 1/3 expansion of the facilities at Strawberry Peak, a new camp a the mouth of Blue Jay Creek, a new camp on the bluffs above the Lake Arrowhead dam, a new camp on crest roads, improvements to many other campgrounds and major road improvements.

In 1927, over 20 acres of camping was established in Barton Flats which had been designated for that purpose by the Forest Service in 1914. By 1938, Barton Flats had become the leading campground complex in Southern California, including five Boy Scout camps, five YMCA, one YWCA, one Camp Fire Girls, and one Church of the Brethren. San Bernardino County leased 560 acres of Barton Flats for a county playground in 1928.

The greatest amount of road, trail and fire break construction occurred during the depression with the help of the unemployed from San Bernardino County and the CCC. From 1933 through 1934, 140 miles of roads, 320 miles of fire breaks, 35 campgrounds, and 33 buildings were constructed or improved. From 1936 through 1938, the CCC added the Miller Canyon public campground and completed the dam at Huston Flat which formed Lake Gregory.

### 1940-present

Recreation activity was somewhat reduced in the mountains during World War II. Servicemen used the resort for "rest and recuperation". But after the war, recreation became a specialization of the Forest Service. Today the mountains are dotted with many private and public campgrounds and their popularity is continuing to grow.

## Locational criteria

### Study area map

As man continues to modify and shape land for various activities, primary and secondary impacts inevitably occur. This landscape change, whether major or minor, affects not only environmental processes, but also economic, socio-cultural, and political processes which will, in turn, affect landscape change. The criteria developed for this project was influenced by H.T. Odum's work (1983) which describes natural systems as flows of energy and materials occurring through different levels of the biosphere. These levels are the lithosphere (soil and subsoil), terisphere/hydrosphere (top of soil to upper plant canopies), and the atmosphere (air region which will sustain life). The major systems of water, energy and nutrients (nitrogen is used as the primary example) are employed to compare the natural system flows without human impact to those which include human impacts. These comparisons are graphically represented in Appendix A.

Based on Odum's approach, the project team decided that the models in this study would be macro-based at a local scale and deal with human time scale effects. Consequently, the following criteria was developed for the locational and operational requirements for the Eco-camp.

### Water

A major natural resource used at the Eco-camp will be water. Water sources for drinking in the San Bernardino Mountains are the lakes and pumped ground water. The water used at the Eco-camp should be the minimum amount necessary, used as efficiently as possible. This can be done by using low volume fixtures. recycling and reusing the available water whenever possible. To reduce the amount of water utilized by conventional toilets, composting toilets will be used. Also, greywater from showers, washing and cooking will be captured and recycled in appropriate ways. Foodstuffs to be grown and eaten onsite will be of the type that adapt to low volume, direct soil irrigation to minimize air-bourne drift and evaporation.

Any site improvement, particularly ones that create impervious surfaces will dramatically impact the water cycle. Consequently, hard surface road construction will be kept to a minimum. Runoff from structures or disturbed areas should be detained and/or captured for reuse. Under no circumstances should concentrated runoff be allowed to accumulate. This will prohibit sedimentation and possible debris flows. The siting of structures and facilities must avoid any natural drainage courses.

Depending on the soil and drainage configurations, some types of grey water could be introduced into existing intermittent streams, ground water, or other water drainage area to encourage the growth of riparian flora to attract local fauna.

### Energy and nutrients

Energy sources for the camp should be non-polluting and renewable, except for emergency situations. Natural vegetation should be preserved to provide food and habitat for local organisms. The forest litter at the soil surface should not be removed except for activity areas and circulation routes. This is especially important in an ecosystem where the decomposer organisms play a crucial role in energy and nutrient cycles. Food wastes from either imported sources or grown onsite should be composted to provide mulch for the camp garden. Extra composted mulch may be introduced into areas impacted by human activity to assist in soil stabilization and restoration.

## Resource data base

A listing of the digital library of maps used in the project is included in Appendix B. The data covers a region of the San Bernardino Mountains from the Cajon Pass to the west to the town of Running Springs to the east and to the foot of the mountains to the north and south. The data, which encompasses approximately 170 square miles, is located within the U.S. Forest Service boundaries contained within the four U.S.G.S. quadrangles of Lake Arrowhead, Silverwood Lake, San Bernardino North and Harrison Mountain.

## Locational model

### Goals

The goals for the Eco-camp locational model include the following characteristics

- sites that are adjacent to the greatest range of natural environments
- sites that will sustain the least negative impact on the natural environment by camp facilities
- sites that can demonstrate efficient use of solar energy
- sites that can provide effective reintroduction of filtered waste water into the environment

**Variables considered**

In support of the above goals the following variables were considered

- within the San Bernardino National Forest and outside the existing urban and state park areas.
- close proximity to urban-historic areas
- between 1/2 and 1/4 mile of perennial streams
- vegetation type
- aspect
- soil erosion factors
- reasonably close proximity to existing roads
- away from existing major power lines

**Map 1, perennial streams, buffer zones, forest vegetation, paved roads**

The design team limited the investigation to the north half of the study area. This approach was taken to locate the camp as far as possible from the densely populated urban areas of the San Bernardino Valley.

Since riparian areas provide habitat for the greatest number of animal and plant species, locations near perennial streams were sought to provide the user with a high diversity of ecosystems. In order to locate the camp within walking distance of a perennial stream yet minimize the camp's impact, the dark *exclusionary* zone, 1/4 mile wide on each side of perennial streams was created. Biologists often use 500 feet as a rule-of-thumb for buffer width, but the additional distance is justified in this model because of the rarity of these exceptionally high valued habitats. An *inclusionary* zone was created that extends 1/2 mile from the edge of the exclusionary zone.

The conifer and broadleaf forests were selected and are shown shaded. The balance of the vegetation is mostly chaparral and was avoided due to high fire hazard. Paved roads are included to indicate their proximity to possible sites.

**Map 2, aspect, vegetation, moderate erosion potential**

South, southeast and southwest aspects were selected to provide the best exposure for solar energy.

Most soils in this national forest are highly erosive. Since no soils are rated low for erosivity, soils rated moderately erosive were selected.

The location of intermittent streams is also shown. Locating the camp close to intermittent streams serves a two-fold purpose. It can (1) enhance habitat value by augmenting stream flow to a level that more closely approaches perennial status, and (2) recycle waste water through a greater range of uses before it percolates to groundwater. An *exclusionary* zone of 1/8 mile wide on each side of intermittent streams is recommended but is not shown.

The proximity to major power transmission lines is considered negative in the locational model and those facilities are indicated. Proximity to existing trails, not included on the map, is considered a positive spatial value.

**Map 3, alternative sites**

Sites which met all the locational goals of the model were identified visually. Site 5 (Rock Camp) was included in the evaluation since it is a well-known area of rich cultural and historic value.

## Evaluation of alternative sites

The following table illustrates the Decision Element Analysis Method used to evaluate the seven alternative sites. Criteria and the rationale for its inclusion in the analysis is presented below.

**Criteria**

- moderate proximity to perennial streams
- moderate soil erosion potential
- lack of proximity to major power lines
- lack of proximity to off road vehicle routes
- close proximity to urban-historic features
- close proximity to existing roads

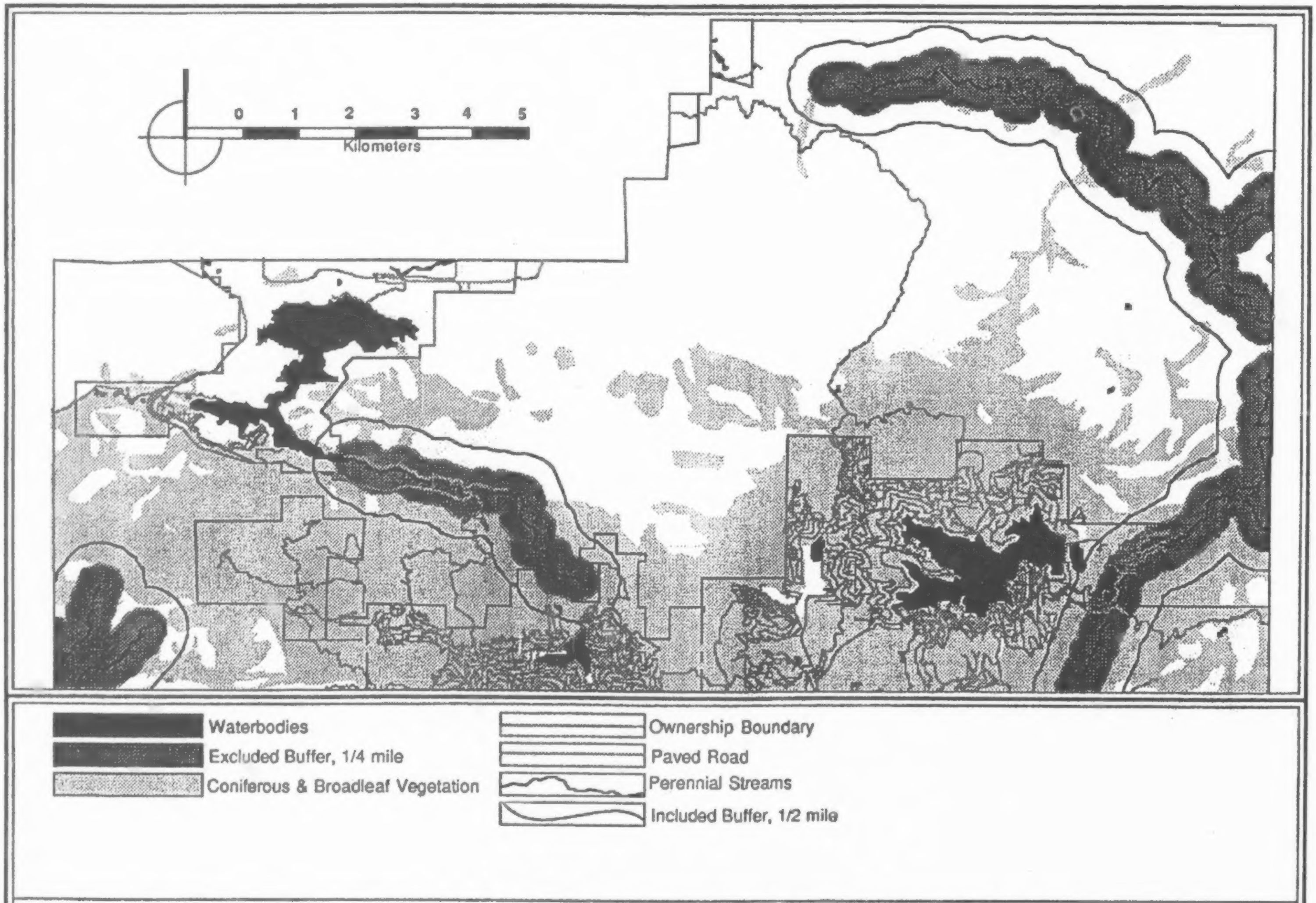
**Rationale**

- provides educational opportunities
- minimizes erosion and siltation
- minimizes possible negative visual and health impacts
- minimizes negative sensory impact
- provides educational opportunities
- minimizes erosion, conserve resources

The numbered site alternatives are listed on the left side of the decision table and evaluative criteria are listed across the top of the table. Each site was rated on each criterion as follows

- 5 points - completely meets criteria
- 3 points - moderately meets criteria
- 0 points - does not meet criteria

Each criterion has a weighting value based on its relative importance to the site locational goals. This value is listed along the bottom row of the table. The criterion rating is multiplied by the weighting value, producing a numerical value for each criterion for each alternative site. Scores are added horizontally to create a total point rating for each site. Those sites receiving the highest overall scores become the preferred alternatives. As the

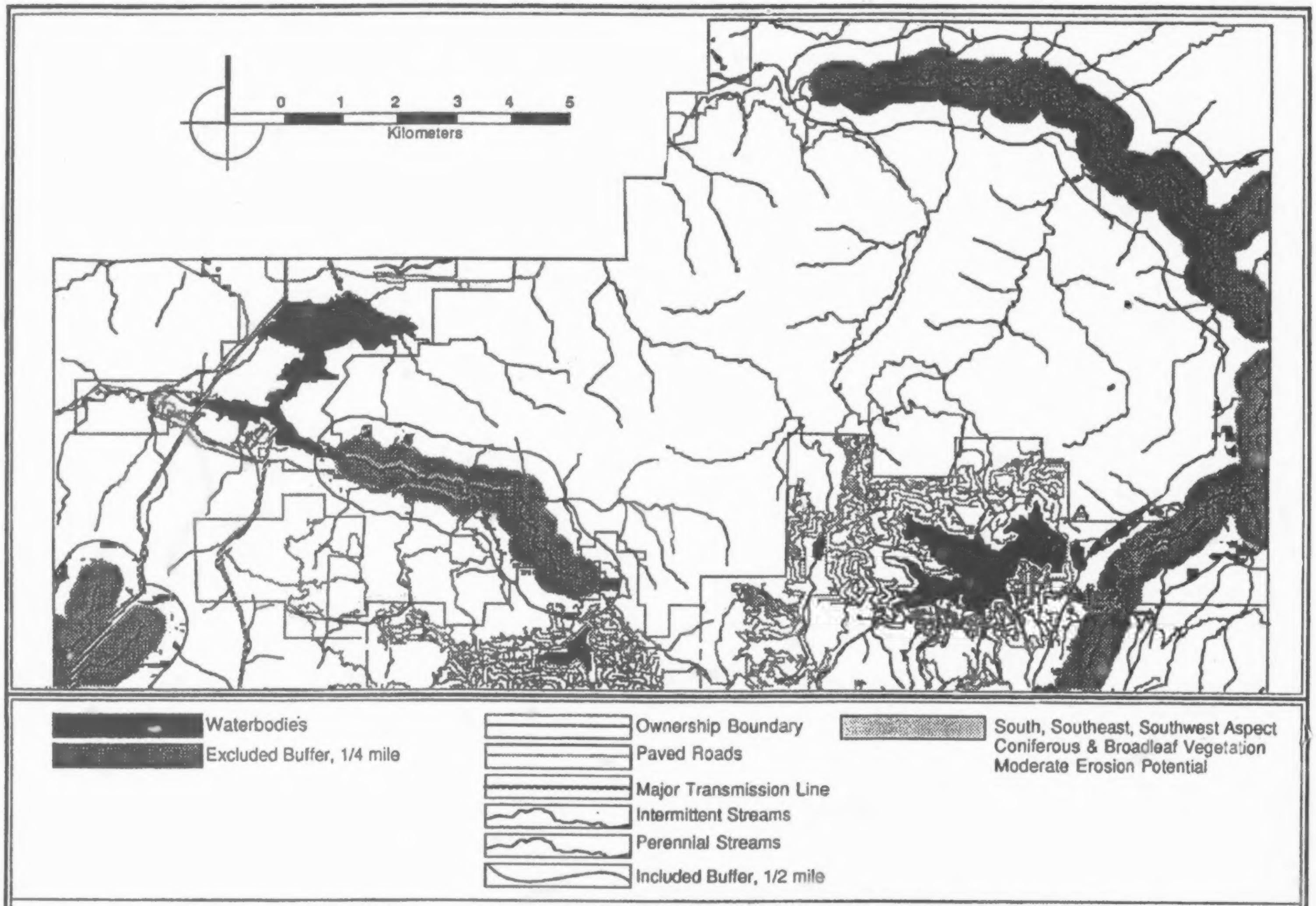

0
1
2
3
4
5
Kilometers
Waterbodies
Excluded Buffer, 1/4 mile
Coniferous & Broadleaf Vegetation
Ownership Boundary
Paved Road
Perennial Streams
Included Buffer, 1/2 mile

ECO-CAMP

**Map 1, Perennial Streams, Buffer Zones, Forest Vegetation, Paved Roads**

San Bernardino National Forest, San Bernardino County, California

**California State Polytechnic University, Department of Landscape Architecture, Pomona, California**

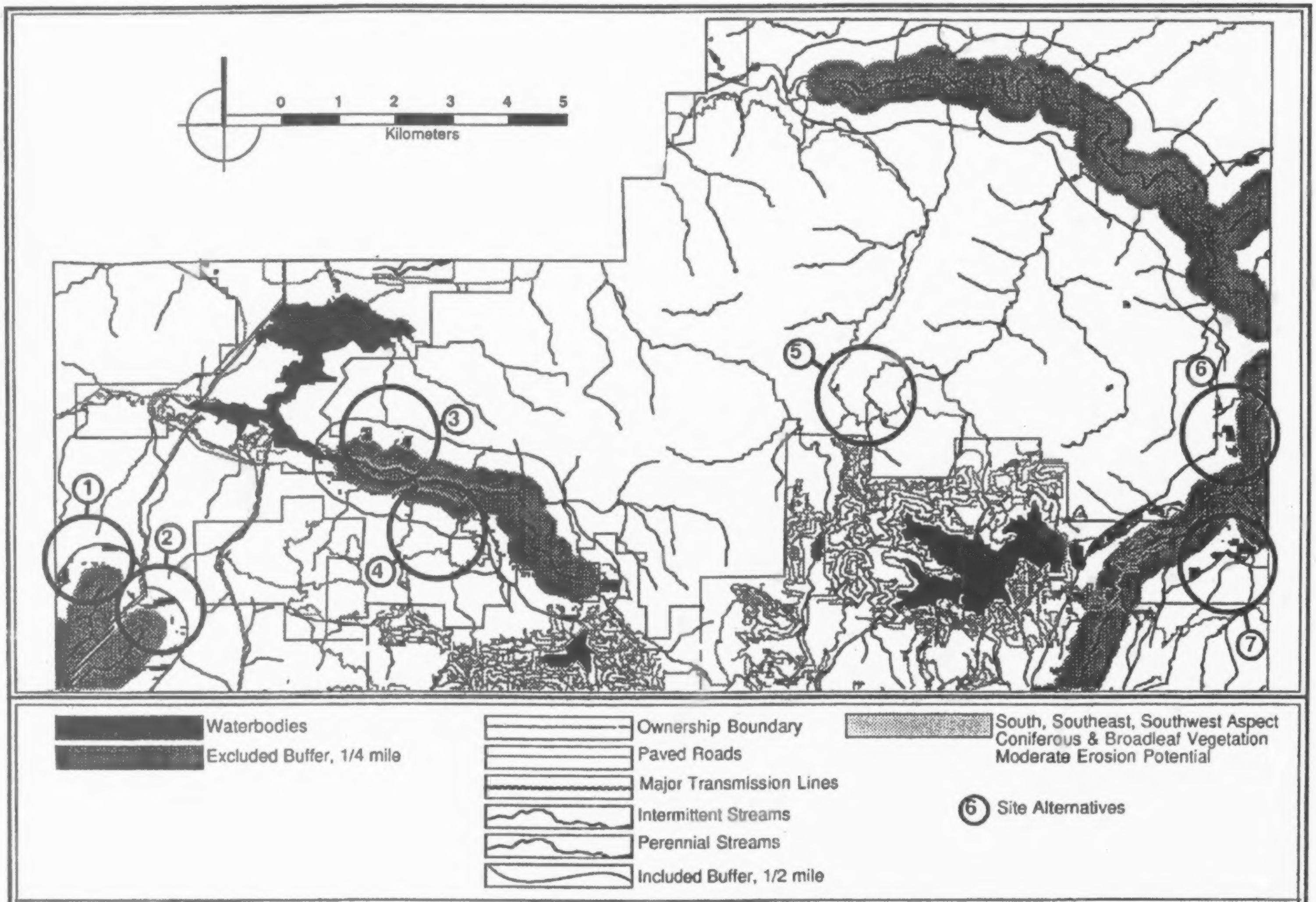

0
1
2
3
4
5
Kilometers
Waterbodies
Excluded Buffer, 1/4 mile
Ownership Boundary
Paved Roads
Major Transmission Line
Intermittent Streams
Perennial Streams
Included Buffer, 1/2 mile
South, Southeast, Southwest Aspect
Coniferous & Broadleaf Vegetation
Moderate Erosion Potential

**ECO-CAMP**

**Map 2, Aspect, Vegetation, Moderate Erosion Potential**

San Bernardino National Forest, San Bernardino County, California

**California State Polytechnic University, Departmenct of Landscape Architecture, Pomona, California**


0
1
2
3
4
5
Kilometers
①
②
③
④
⑤
⑥
⑦
Waterbodies
Excluded Buffer, 1/4 mile
Ownership Boundary
Paved Roads
Major Transmission Lines
Intermittent Streams
Perennial Streams
Included Buffer, 1/2 mile
South, Southeast, Southwest Aspect
Coniferous & Broadleaf Vegetation
Moderate Erosion Potential
⑥ Site Alternatives

**Map 3, Alternative Sites**

ECO-CAMP

San Bernardino National Forest, San Bernardino County, California

**California State Polytechnic University, Department of Landscape Architecture, Pomona, California**

table indicates, alternatives four and seven appear to be the most favorable sites for the project.

| Criteria<br>Alternatives | Close proximity to perennial streams | Mod. soil erosion potential | Lack of proximity to major power lines | Lack of proximity to off-road vehicles | Close proximity to urban-historical | Close proximity to existing roads | weighted score |
|---|---|---|---|---|---|---|---|
| Site #1 | ● | ○ | ◍ | ◍ | ○ | ◍ | 61 |
| Site #2 | ● | ○ | ◍ | ◍ | ○ | ◍ | 61 |
| Site #3 | ● | ○ | ● | ◍ | ○ | ● | 77 |
| ▶ Site #4 | ● | ◍ | ● | ◍ | ○ | ● | 92 |
| Site #5 | ○ | ○ | ● | ○ | ● | ● | 55 |
| Site #6 | ● | ◍ | ● | ○ | ○ | ○ | 65 |
| ▶ Site #7 | ● | ◍ | ● | ● | ○ | ◍ | 94 |
| Criteria Weighting | 5 | 5 | 5 | 4 | 3 | 3 | |

## Conclusions

The process used and the criteria applied in this project seemed to yield reasonable results. The next stage of this study should test the accuracy and credibility of the above sites. A quick field observation of the preferred sites seemed to indicate a good 'match' between the criteria and the chosen sites. But, the findings of our analysis at this scale has to be further verified. For example, the size of each of the the preferred sites varies, consequently further study at a smaller scale is needed to determine the results after further elaboration of the program and the camp design.

The criteria used in the evaluation could be challenged. For example, how would the study be altered if only south aspects or if different vegetative families were chosen? What would be the result if different weightings were given to other criteria?

The data base used for this project was designed and developed before the modelling was performed. A more appropriate sequence would be to let the criteria define the data base rather than the other way. Therefore, one might legitimately ask what additional criteria might be required for this study and/or, was the data used of sufficient detail to provide adequate results.

In the ten weeks allowed for this study, the study team was able to test many hypothesis interactively, a capacity of GIS that greatly enhances the design process.

## Bibliography

Bailey, Paul. "Walkara. Hawk of the Mountains", Westernlore Press, Los Angeles, 1954.

Clark, Anthony Orr. "Quaternary Evolution of the San Bernardino Valley." Quarterly of the San Bernardino County Museum Association, Vol. XXVI, No. 2 and 3, Winter, 1978, Spring 1979.

Frazee, W.D. "San Bernardino County, Its Climate and Resources", San Bernardino Daily and Weekly Argus, Office of San Bernardino California. 1876.

"Historical Landmarks of San Bernardino County", Quarterly of the San Bernardino County Museum Association, Vol. XXVIII, No. 1 and 2, Fall 1980.

Holcomb, W.F. "Reminiscences and History of Holcomb Valley", W.F. Holcomb, 1900.

LaFuze, Pauliena B. "Saga of the San Bernardinos", Volume I. San Bernardino County Museum Association, 1971

LaFuze, Pauliena B. "Saga of the San Bernardinos", Volume II. San Bernardino County Museum Association, 1971.

Larsen K., Hulse D. MacGIS, University of Oregon, Eugene Oregon. 1989.

Leadabrand, Russ. "Guidebook to the San Bernardino Mountains of California." The Ward Ritchie Press, Los Angeles. 1964.

Lyle, J.T. "Design for Human Ecosystems." Van Nostrand Reinhold Co., New York. 1985.

Odum, H.T. "Systems Ecology An Introduction." Wiley, New York. 1983.

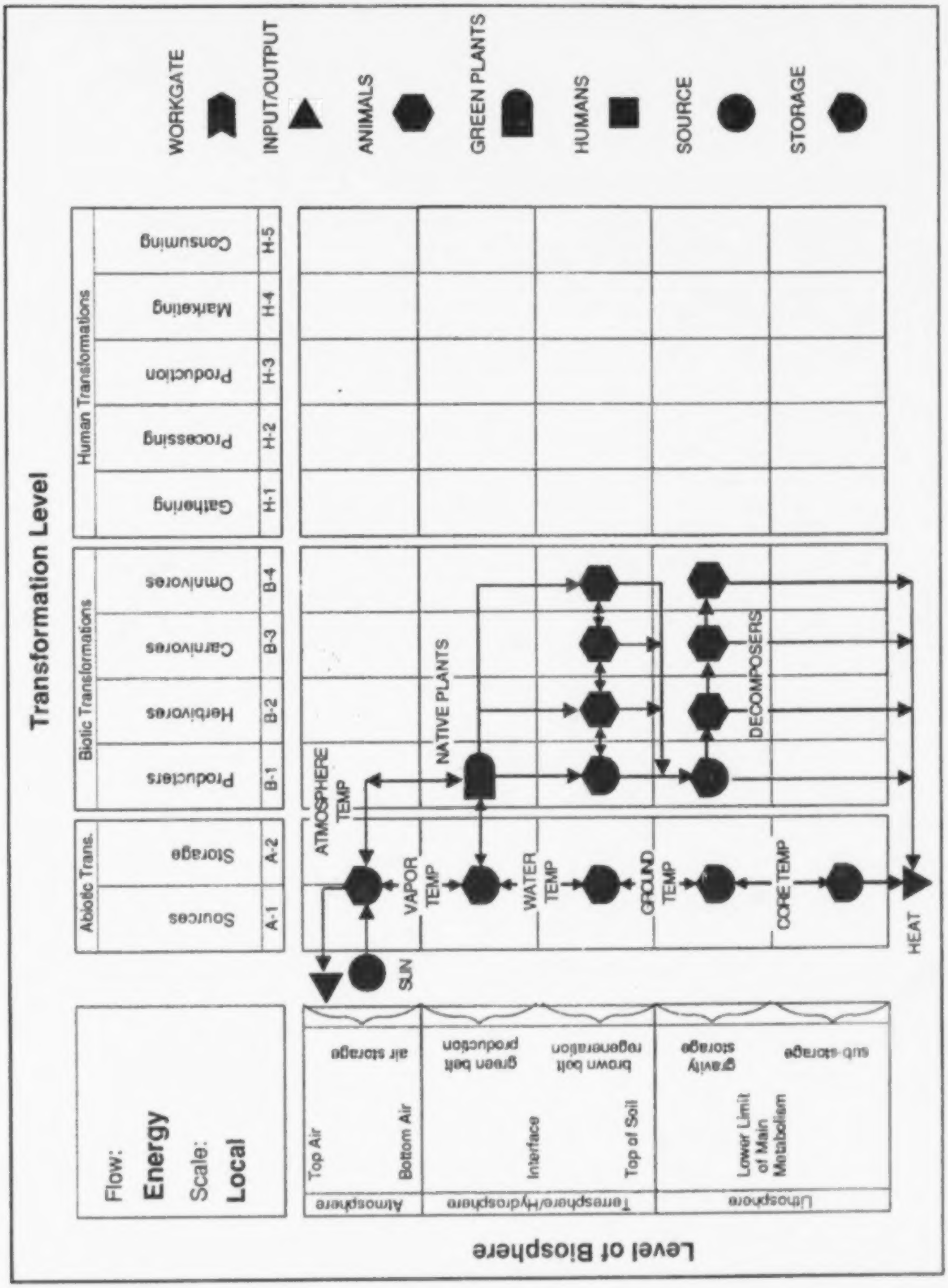
Robinson, John W. "San Gorgonio Mountain The Early Years." San Bernardino County Museum Association. Vol. XXIII. No. 2 and 3. Winter and Spring 1976.

Robinson, John W. "San Bernardino Mountain Trails." Wilderness Trails, Berkeley. 1979.

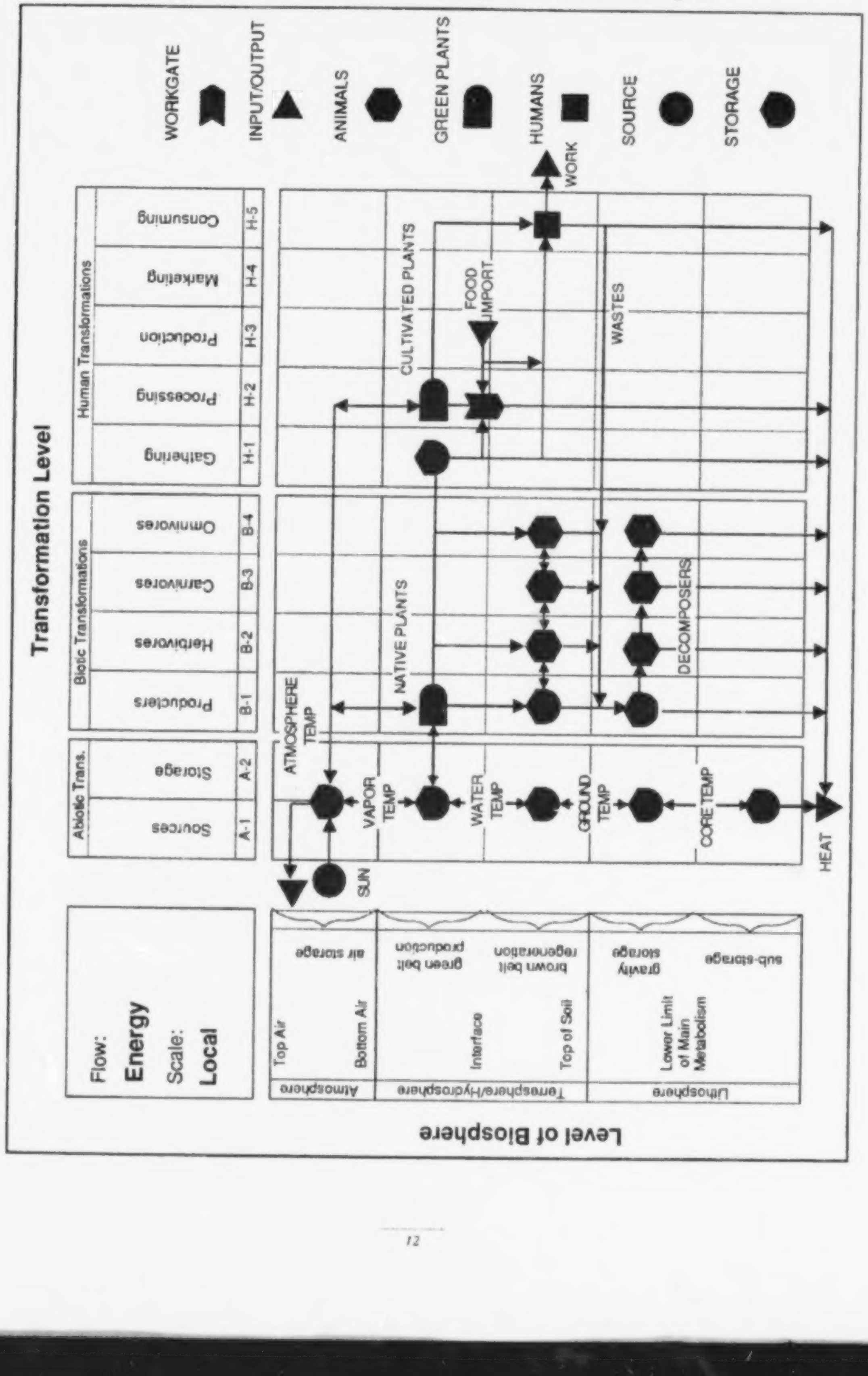
Robinson, John W. "The San Bernardinos." Big Santa Anita Historical Society, 1989

Stone, Kendall J. "Foxsong-100 years of cow ranching in the San Bernardino Mountains and Mojave Desert." Sagebrush Press, 1989.

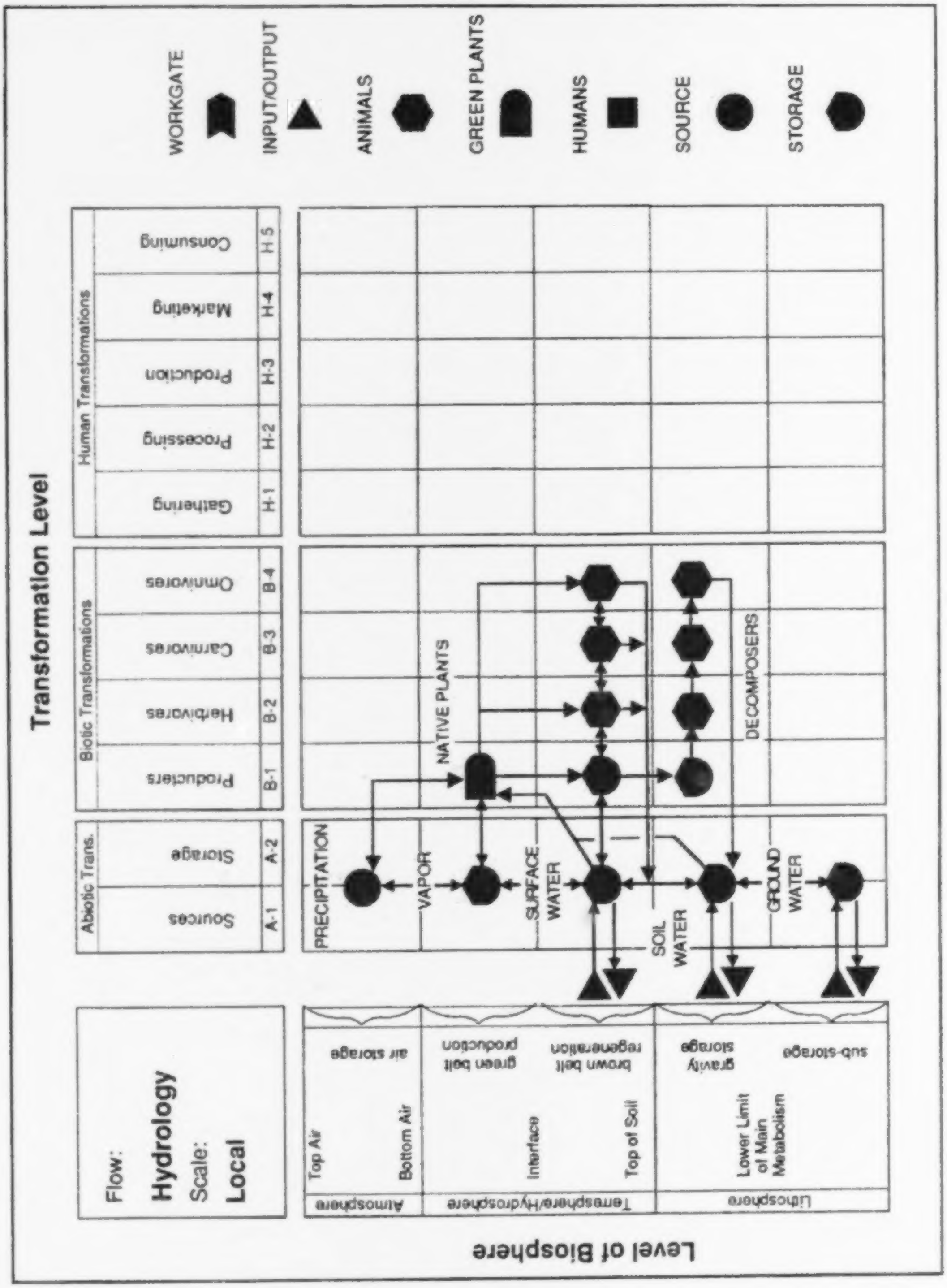
## Appendix A, natural systems energy flow


Transformation Level
Abiotic Trans.
Sources
A-1
Storage
A-2
Biotic Transformations
Producers
B-1
Herbivores
B-2
Carnivores
B-3
Omnivores
B-4
Human Transformations
Gathering
H-1
Processing
H-2
Production
H-3
Marketing
H-4
Consuming
H-5
WORKGATE
INPUT/OUTPUT
ANIMALS
GREEN PLANTS
HUMANS
SOURCE
STORAGE
Flow:
Energy
Scale:
Local
Level of Biosphere
Atmosphere
Top Air
Bottom Air
air storage
Terresphere/Hydrosphere
Interface
Top of Soil
green belt production
brown belt regeneration
Lithosphere
Lower Limit of Main Metabolism
gravity storage
sub-storage
SUN
ATMOSPHERE TEMP
VAPOR TEMP
WATER TEMP
GROUND TEMP
CORE TEMP
HEAT
NATIVE PLANTS
DECOMPOSERS

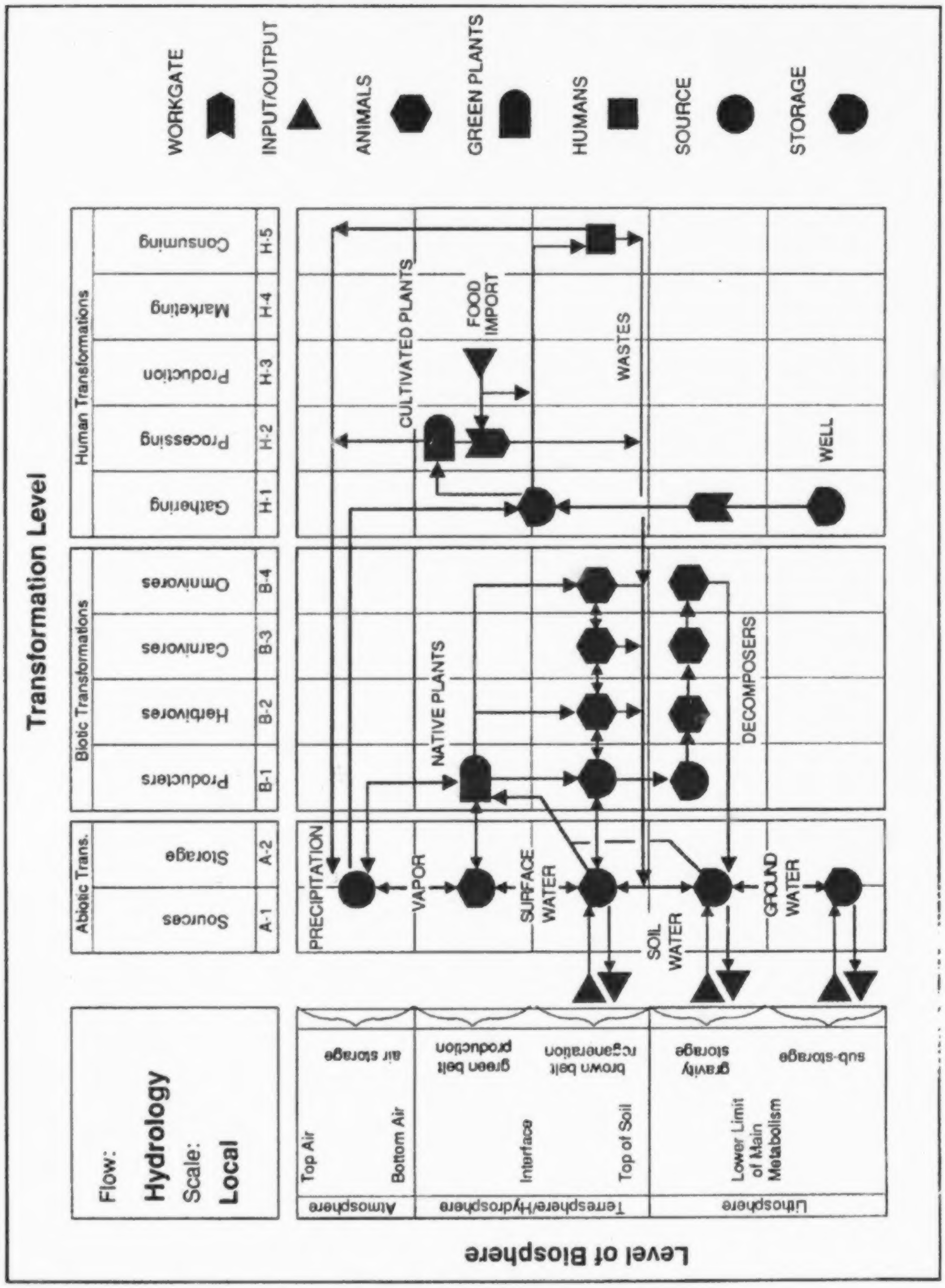
## Appendix A, natural systems energy flow w/ human impact


Transformation Level
Abiotic Trans.
Sources
Storage
A-1
A-2
Biotic Transformations
Producers
Herbivores
Carnivores
Omnivores
B-1
B-2
B-3
B-4
Human Transformations
Gathering
Processing
Production
Marketing
Consuming
H-1
H-2
H-3
H-4
H-5
WORKGATE
INPUT/OUTPUT
ANIMALS
GREEN PLANTS
HUMANS
SOURCE
STORAGE
WORK
SUN
ATMOSPHERE TEMP
VAPOR TEMP
WATER TEMP
GROUND TEMP
CORE TEMP
HEAT
NATIVE PLANTS
CULTIVATED PLANTS
FOOD IMPORT
WASTES
DECOMPOSERS
Flow:
Energy
Scale:
Local
Atmosphere
Terrasphere/Hydrosphere
Lithosphere
Top Air
Bottom Air
Interface
Top of Soil
Lower Limit of Main Metabolism
air storage
green belt production
brown belt regeneration
gravity storage
sub-storage
Level of Biosphere

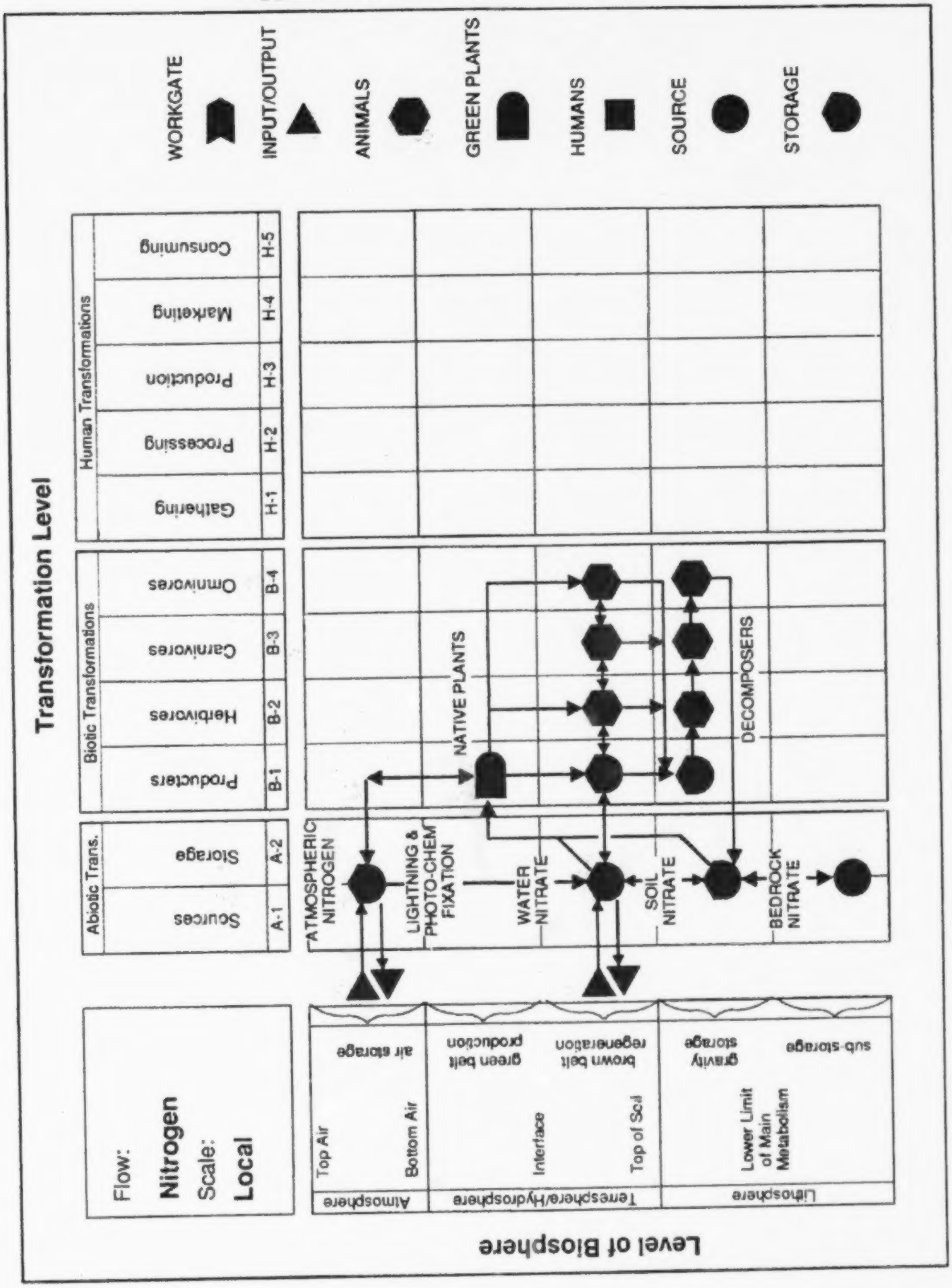
## Appendix A, natural systems water flow


Transformation Level
Abiotic Trans.
Sources
A-1
Storage
A-2
Biotic Transformations
Producers
B-1
Herbivores
B-2
Carnivores
B-3
Omnivores
B-4
Human Transformations
Gathering
H-1
Processing
H-2
Production
H-3
Marketing
H-4
Consuming
H-5
PRECIPITATION
VAPOR
SURFACE WATER
SOIL WATER
GROUND WATER
NATIVE PLANTS
DECOMPOSERS
WORKGATE
INPUT/OUTPUT
ANIMALS
GREEN PLANTS
HUMANS
SOURCE
STORAGE
Flow:
Hydrology
Scale:
Local
Atmosphere
Terrasphere/Hydrosphere
Lithosphere
Top Air
Bottom Air
Interface
Top of Soil
Lower Limit of Main Metabolism
air storage
green belt production
brown belt regeneration
gravity storage
sub-storage
Level of Biosphere

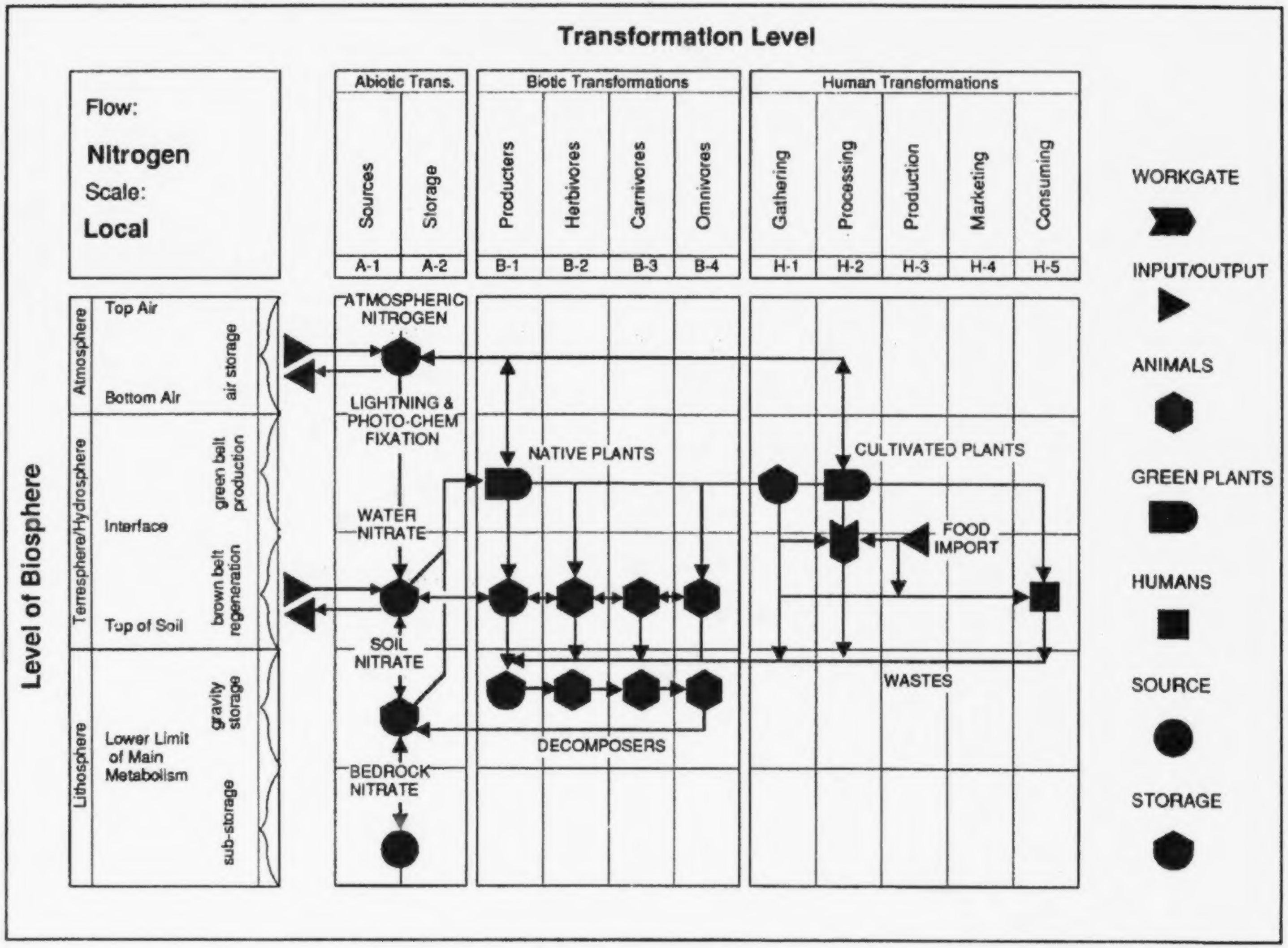
## Appendix A, natural systems water flow w/ human contact


Transformation Level
Abiotic Trans.
Sources
A-1
Storage
A-2
Biotic Transformations
Producers
B-1
Herbivores
B-2
Carnivores
B-3
Omnivores
B-4
Human Transformations
Gathering
H-1
Processing
H-2
Production
H-3
Marketing
H-4
Consuming
H-5
WORKGATE
INPUT/OUTPUT
ANIMALS
GREEN PLANTS
HUMANS
SOURCE
STORAGE
PRECIPITATION
VAPOR
SURFACE WATER
SOIL WATER
GROUND WATER
NATIVE PLANTS
DECOMPOSERS
CULTIVATED PLANTS
FOOD IMPORT
WASTES
WELL
Flow:
Hydrology
Scale:
Local
Atmosphere
Top Air
Bottom Air
Terresphere/Hydrosphere
Interface
Top of Soil
Lithosphere
Lower Limit of Main Metabolism
air storage
green belt production
brown belt regeneration
gravity storage
sub-storage
Level of Biosphere

## Appendix A, natural systems nutrient flow


Transformation Level
Abiotic Trans.
Sources
A-1
Storage
A-2
Biotic Transformations
Producers
B-1
Herbivores
B-2
Carnivores
B-3
Omnivores
B-4
Human Transformations
Gathering
H-1
Processing
H-2
Production
H-3
Marketing
H-4
Consuming
H-5
WORKGATE
INPUT/OUTPUT
ANIMALS
GREEN PLANTS
HUMANS
SOURCE
STORAGE
Flow:
Nitrogen
Scale:
Local
Level of Biosphere
Atmosphere
Top Air
Bottom Air
air storage
Terresphere/Hydrosphere
Interface
Top of Soil
green belt production
brown belt regeneration
Lithosphere
Lower Limit of Main Metabolism
gravity storage
sub-storage
ATMOSPHERIC NITROGEN
LIGHTNING & PHOTO-CHEM FIXATION
WATER NITRATE
SOIL NITRATE
BEDROCK NITRATE
NATIVE PLANTS
DECOMPOSERS

## Appendix A, natural systems nutrient flow w/ human contact


Transformation Level
Flow:
Nitrogen
Scale:
Local
Abiotic Trans.
Sources
Storage
A-1
A-2
Biotic Transformations
Producters
Herbivores
Carnivores
Omnivores
B-1
B-2
B-3
B-4
Human Transformations
Gathering
Processing
Production
Marketing
Consuming
H-1
H-2
H-3
H-4
H-5
WORKGATE
INPUT/OUTPUT
ANIMALS
GREEN PLANTS
HUMANS
SOURCE
STORAGE
Level of Biosphere
Atmosphere
Top Air
Bottom Air
air storage
Terresphere/Hydrosphere
Interface
Top of Soil
green belt production
brown belt regeneration
Lithosphere
Lower Limit of Main Metabolism
gravity storage
sub-storage
ATMOSPHERIC NITROGEN
LIGHTNING & PHOTO-CHEM FIXATION
WATER NITRATE
SOIL NITRATE
BEDROCK NITRATE
NATIVE PLANTS
CULTIVATED PLANTS
FOOD IMPORT
WASTES
DECOMPOSERS

## Appendix B, resource data base

1. Hydrology (manually digitized from USGS, 124,000)
    - Primary streams
    - Secondary streams
    - Intermittent streams
    - Water bodies
2. Watersheds (manually digitized from interpreted USGS, 124,000)
    - Watershed boundaries
3. Vegetation (manually digitized from reduced 124,000 from USFS)
    - 76 data categories
4. Roads (manually digitized from USGS, 124,000)
    - Heavy duty
    - Medium duty
    - Light duty
    - Unimproved
5. Soil (manually digitized from reduced 124,000 soil maps from USFS)
    - 69 data categories
    - Hydrologic groups
    - Erosion Hazard ratings
6. Power (manually digitized from USGS, 124,000)
    - Major power lines
7. Elevation (manually digitized from USGS, 124,000)
    - <3000'
    - 3000-4000'
    - 4000-5000'
    - 5000-6000'
    - >6000'
8. Trails (manually digitized from USGS, 124,000)
    - Trails and firebreaks
9. Square Mile Sections (manually digitized from USGS, 124,000)
    - 1 square mile section grid
10. Border (manually digitized from USGS, 124,000)
    - Forest Service boundary
11. Aqueduct (manually digitized from USGS, 124,000)
    - The California Aqueduct
12. Landuse (manually digitized from various sources)
    - Land Use Features
13. Water (manually digitized from USGS, 124,000)
    - water bodies
14. Focal (manually digitized from various sources)
    - Landscape features
15. Aspect (USGS, 124,000 DEM, processed in MacGIS (Larsen Hulse 1989))
    - north, northeast, east, southeast, south, southwest, west, northwest, flat
16. Slope (USGS, 124,000 DEM, processed in MacGIS (Larsen Hulse 1989))
    - 0-5%, 6-10%, 11-15%, 16-20%, 21-25%, 26-30%, 31-40%, 41-50%, 51-60%, >60%

# Plenary presentations

---

**A DEM (DTM) of the GIS navel**

John Keene, Genasys II Pty., North Sydney, NSW, Australia

**Working smarter amidst chaos**

Hans Schreier et al., Resource Management Science, University of British Columbia, Vancouver,BC, Canada

**Smarter applications — where do we go from here?**

Joseph K. Berry, GIS World Inc., Fort Collins, CO, USA

**New technologies for creating public windows into GISs**

Jack Dangermond, Environmental Systems Research Institute, Redlands, CA, USA

*** Environmentally smarter working — fact or fancy?**

Ellsworth LeDrew, Earth Observations Laboratory, Institute for Space and Terrestrial Science, University of Waterloo, Waterloo, ON, Canada

*** Are we really working smarter in our forests — or just faster ?**

David Goodenough, Forestry Canada, Pacific Forestry Centre, Victoria, BC, Canada

*** Technologies and methodologies — are better solutions emerging ?**

Michael F. Goodchild, National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA, USA

*** The corporate database — making the marriage work**

Stephen Guptill, US Geological Survey, Reston, VA, USA

** Abstracts not available at time of printing*

**GIS'92**
**Symposium**

Vancouver
British Columbia
February 1992

# A DEM (DTM) of the GIS navel

**John Keene**

Genasys II Pty Ltd.
13th Level 33 Berry Street
North Sydney, NSW 2060
Tel: 02-954-0022 Fax: 02-954-9930

## Abstract

*GIS is a very introspective industry. Geographers are more focussed on doing what they are good at rather than what needs to be done. Product is produced without a focussed view of market direction or general requirements. Too little emphasis is placed on data quality, location, or query-ability — and on real projects — too much emphasis on consulting, gee-whizery, pilot projects and GIS-ing about.*

*This paper is an introspective look at the GIS industry . It will answer such questions as:*

*Why is UNIX pronounced* ***Eunuchs****?*

*Why do GIS "experts" approach intellectual challenges with an open mouth?*

*How do we resolve the D3 dilemma — Decidedly Dodgy Data?*

*Why is GIS a reality distortion field?*

*The paper is planned to be of special interest to anyone (user or supplier) who actually wants to get something done!*

GIS
92

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# Working smarter amidst chaos

**Hans Schreier**
**Sandra Brown**

Resource Management Science
University of British Columbia,
Vancouver, B.C.

## Abstract

*Recent publications have suggested that the GIS industry is in a state of chaos. According to many authorities, the environment is also in a state of chaos. So, does* (GIS chaos) + (environmental chaos) = (catastrophe), *or can we develop smart systems that bring some semblance of control and direction into uncertainty at the local, national and global scales?*

*To find smart solutions in GIS-based environmental evaluations the four key components of hardware, software, data and people need to have a symbiotic relationship. Hardware developments are somewhat predictable and although they are happening at an increasing pace, efforts are under way to provide compatibility and continuity. This is not so in the software realm where translators, compatibility and new developments may be as unpredictable as a pit bull terrier. The disorder and turmoil of spatially-referenced environmental data is likely the most problematic factor waiting for urgent solutions. This is especially acute when trying to examine environmental data in a holistic manner. Finally, an environmentally successful GIS system has to be piloted by people that have good grounding in resources, database management and computers.*

***Working smarter** means you must select and collect appropriate data, use inexpensive hardware, flexible and compatible software, and choose bright and enthusiastic resource people with a computer slant. Choose your project area carefully to avoid spending 95% of your time building, correcting and updating your database. Add models to your evaluation that work outside of GIS but can be easily integrated. Then integrate imagery, CAD, animated graphics and GPS in conjunction with GIS analysis.*

*Examples of such GIS integration will be provided for a number of environmental evaluations covering applications in resource conflict resolution, water pollution and management, forest degradation, biomass sustainability evaluation, environmental pollution and quantifying land use dynamics. The examples cover local, national and international examples ranging from sea level to the tallest mountains in the Himalayas.*

GIS'92
Symposium

Vancouver
British Columbia
February 1992

# Smarter applications — where do we go from here?

**Joseph K. Berry**

GIS World Inc.
Suite 280 2629 Redwing Road
Fort Collins, CO USA 80526
Tel: 303-223-4848 Fax: 303-223-5700

## Abstract

*GIS technology is a technical evolution that has become a revolution. Applications in the 1970s and 80s moved from computer mapping to spatial database management primarily in response to more efficient operations in data storage, retrieval and geographic query. Such applications are primarily descriptive of landscape features and facilities. More recently, applications have moved to cartographic and spatial modeling involving prescriptive interpretations of mapped data. Today, for example, GIS not only describes an existing set of roads, but contributes to the analysis to locate a new road. In such applications, the emphasis moves from data composites in graphic form for human viewing to map variable interrelationships expressed through quantitative models — a sort of "map-ematics".*

*The expanding role of GIS requires users to rethink their understanding of map content, structure and use. Entirely new "tools," such as spatial statistics, error propagation, visual exposure density and weighted distance, sharply contrast with traditional map analysis procedures utilizing pens, planimeters, shadings and acetate sheets. GIS is no longer merely the automation of existing map analysis techniques and decision models, but the development of entirely new procedures which takes advantage of new analytical capabilities. The technology's future, however, is more than mapping, management and modeling. GIS is increasingly recognized as a Decision Support System (DSS) which is radically changing the decision-making environment itself, with far-reaching social and legal ramifications. In the 1990s, GIS will become less a technology and more an entirely new way of thinking and communicating through smarter applications.*

**GIS'92**

**Symposium**

Vancouver
British Columbia
February 1992

# New technologies for creating public windows into GISs

**Jack Dangermond**

Environmental Systems Research Institute
380 New York Street
Redlands, CA 92373, USA
Tel: 714-793-2853 Fax: 714-793-5953

## Abstract

*Current and near-future developments in computer and communications hardware and in GIS software are providing the means for creating user-friendly public windows into GISs which have been accessible only by persons with both minimum training in GIS use and the technical means of accessing the GIS's information. Fiber optic networks of several kinds, CD-ROM stored multi-media information, increasingly inexpensive personal computers (or even new video game systems), high definition television, publication of GIS data in inexpensive forms, and the availability of inexpensive GIS information access software which requires no user training, are some of these developments. Prototypes of such public window systems are described and illustrated. The implications of this major development, for the GIS field, for government and private agencies which maintain GIS information, and for the public which is about to become the chief user of GIS technology, are addressed.*